Wolf-Rüdiger Recksiegel

Die Anwendung der Regressions- und Korrelationsanalyse in der Kostenrechnung

1971

1972

DIE ANWENDUNG DER REGRESSIONS- UND KORRELATIONSANALYSE IN DER KOSTENRECHNUNG

INAUGURAL-DISSERTATION
zur Erlangung des
akademischen Grades eines Doktors der Wirtschaftswissenschaft
durch den
Fachbereich Wirtschafts- und Sozialwissenschaften
der Westfälischen Wilhelms-Universität
zu Münster

vorgelegt von
Dipl. Kaufmann Wolf-Rüdiger Recksiegel
aus Bielefeld

1972

UNIV.
LEIDEN
BIBL.

1. Berichterstatter: Prof. Dr. J. Bleymüller
2. Berichterstatter: Prof. Dr. U. Leffson
Dekan: Prof. Dr. J. Bleymüller

Tag der mündlichen Prüfung: 3./4. Februar 1971

INHALTSVERZEICHNIS

Seite

INHALTSVERZEICHNIS I

VERZEICHNIS DER TABELLEN IM TEXT VII

VERZEICHNIS DER ABBILDUNGEN IM TEXT IX

1. Kapitel: EINFÜHRUNG 1

1.1 Problemstellung 1

1.2 Zielsetzung 5

2. Kapitel: GRUNDLAGEN DER REGRESSIONS- UND KORRELATIONSANALYSE 9

2.1 Begriff und Aufgaben 9

2.2 Schätzungen nach der Methode der kleinsten Quadrate 13

2.3 Nichtlineare Regressionsfunktionen 20

2.4 Multiple und partielle Regression 23

2.5 Der Korrelationskoeffizient 27

3. Kapitel: AUFSTELLUNG EINER GESAMTBETRIEBLICHEN KOSTENFUNKTION 31

4. Kapitel: ERFASSUNG UND VERRECHNUNG VON KOSTEN . 37

4.1 Einführung 37

4.2 Erfassung von Gemeinkostenmaterial im Rahmen der Kostenartenrechnung 37

4.3 Zuordnung und Verrechnung von Kosten im Rahmen der Kostenstellen- und Kostenträgerrechnung 44

4.3.1 Ausgangspunkt der Untersuchung 44

4.3.2 Verrechnung von Gemeinkosten mit Hilfe der multiplen Regressionsanalyse 46

4.3.3 Interpretation der Regressionskonstanten 50

Seite

4.3.4 Vergleich der Regressionsergebnisse mit einer herkömmlichen Verrechnungsmethode 55

4.4 Wahl der geeigneten Schlüsselgröße für die Gemeinkostenverrechnung 58

4.4.1 Problembeschreibung 58

4.4.2 Kostenverursachungsprinzip und Proportionalitätprinzip 60

4.4.3 Die empirische Bestimmung der Gemeinkostenfunktion 64

4.4.4 Test des Korrelationskoeffizienten .. 69

4.4.5 Kritik der statistischen Ermittlung der optimalen Schlüsselgröße 74

4.4.5.1 Allgemeine Beurteilung 74

4.4.5.2 KOCHs Kritik am Proportionalitätsprinzip und den darauf aufbauenden statistischen Verfahren 77

4.4.5.3 Zusammenfassung 80

5. Kapitel: ANALYSE VON KOSTEN ZUM ZWECK DER KOSTENPLANUNG 81

5.1 Einführung 81

5.2 Die Ermittlung der Sollkostenkurve (Gemeinkostenplanung) in der Plankostenrechnung .. 82

5.2.1 Das Prinzip der flexiblen Gemeinkostenplanung 82

5.2.2 Die Anwendung der einfachen linearen Regressionsanalyse 84

5.2.3 Beurteilung des statistischen Verfahrens 86

5.2.3.1 Vergangenheitszahlen als Ausgangspunkt der Analyse 86

5.2.3.2 Extrapolation der Regressionsgeraden 89

5.2.3.3 Bedeutung der Regressionsanalysen in der Praxis 91

5.2.3.4 Zusammenfassung 94

Seite

5.3 Aufstellung von Kostenbudgets 95

5.3.1 Begriff der Budgetkosten im Vergleich zu Plankosten 95

5.3.2 Die Bedeutung kurzfristiger Vorschaurechnungen 97

5.3.3 Die schrittweise Regressionsanalyse . 99

5.3.3.1 Darstellung des Verfahrens .. 99

5.3.3.2 Vorteile gegenüber der herkömmlichen Methode 105

5.3.3.3 Die Anwendung der schrittweisen Regressionsanalyse ... 107

5.3.4 Multikollinearität bei multiplen Analysen 111

5.4 Kostenauflösung zur Gewinnung von Unterlagen für kurzfristige Entscheidungen im Produktionsbereich 118

5.4.1 Grundlagen der statistischen Kostenauflösung 119

5.4.2 Sortimentsplanung bei verbundener Produktion 124

5.4.2.1 Problemstellung 124

5.4.2.2 Ermittlung der Grenzkosten durch partielle Regressionskoeffizienten 126

5.4.2.3 Beurteilung der regressionsanalytischen Grenzkostenermittlung 129

5.4.3 "Make-or-buy"-Entscheidungen 131

5.5 Kostenanalysen im weitesten Sinn 136

5.6 Zuverlässigkeit der Schätzungen - Bildung von Konfidenzintervallen 141

5.6.1 Problemstellung 141

5.6.2 Erläuterung des Zufallsmodells 142

5.6.3 Verteilung der Zufallsvariablen 144

5.6.3.1 Die Normalverteilung 145

5.6.3.2 Der zentrale Grenzwertsatz .. 147

5.6.3.3 Die Ungleichung von TSCHEBYSCHEFF 148

Seite

5.6.4 Standardfehler der Regressionskoeffizienten 150

5.6.5 Standardfehler der Schätzung, Standardfehler der Regressionsgeraden und Standardfehler einer individuellen Prognose 155

5.7 Regressionsgleichungen als Grundlagen für ein betriebliches Matrizenmodell 161

5.8 Zusammenfassende Beurteilung der statistischen Kostenanalyse 165

6. Kapitel: KOSTENKONTROLLE 167

6.1 Einführung 167

6.2 Innerbetrieblicher Zeitvergleich 168

6.2.1 Bildung eines Konfidenzintervalls um die Regressionsgerade 168

6.2.2 Aufstellung einer Kostenkontrollkarte 174

6.2.3 Testen der Nullhypothese 180

6.2.4 Ermittlung eines kritischen Wertes für den Fehler erster Art 182

6.2.5 Der "Signaleffekt" der Kontrollkartentechnik 189

6.3 Zwischenbetrieblicher Kostenvergleich 190

6.3.1 Einführung 190

6.3.2 Stückkostenuntersuchung mehrerer Betriebe unter Verwendung nichtlinearer Regressionsfunktionen 191

6.3.3 Die Form der Regressionsfunktion 200

6.3.3.1 Allgemeine Überlegungen..... 200

6.3.3.2 Extrapolation der Regressionsfunktion bei Approximation einer Kurve durch eine Gerade 201

6.3.3.3 Prüfung der Linearität einer Regression 203

6.3.3.3.1 Muster der Residuen 203

6.3.3.3.2 Varianzanalyse 204

6.3.3.4 Zusammenfassung 211

Seite

7. Kapitel: EINIGE BESONDERE PROBLEME DER KOSTENUNTERSUCHUNG MIT HILFE DER REGRESSIONSANALYSE 215

7.1 Autokorrelation bei Zeitreihen 215

7.1.1 Begriff der Zeitreihe 215

7.1.2 Probleme bei Zeitreihen 216

7.1.3 Möglichkeiten zur Feststellung, Verringerung oder Beseitigung der Autokorrelation 217

7.2 Heteroskedasie 223

7.3 Gewinnung des Datenmaterials 225

7.3.1 Länge der Beobachtungszeiträume - sachliche und zeitliche Abgrenzung des Datenmaterials 225

7.3.2 Stichprobenumfang 229

7.3.3 Streuungsintervall der unabhängigen Veränderlichen 231

7.3.4 Meßfehler 232

8. Kapitel: ZUSAMMENFASSENDE INTERPRETATION DER UNTERSUCHUNGSERGEBNISSE IM HINBLICK AUF DIE ZIELSETZUNG 236

9. ANHANG I: RECHNERISCHE DURCHFÜHRUNG VON REGRESSIONS- UND KORRELATIONSANALYSEN 252

9.1 Manuelle Berechnungen 252

9.2 Berechnungen mit programmierbaren Tischrechnern 258

9.3 Berechnungen mit elektronischen Datenverarbeitungsanlagen 260

9.3.1 International Business Machines (IBM-System 360) 261

9.3.2 Siemens 268

9.3.3 Bull General Electric 271

9.3.4 International Computers Limited (ICL) 274

9.3.5 Zuse 275

9.3.6 Deutsches Rechenzentrum 275

Seite

10. ANHANG II: EINIGE WICHTIGE TABELLEN UND NOMOGRAMME 283

Tabelle 1*: Standardnormalverteilung (einseitig) 284

Tabelle 2*: Standardnormalverteilung (zweiseitig) 285

Tabelle 3*: Studentverteilung (einseitig) 286

Tabelle 4*: Studentverteilung (zweiseitig) 287

Tabelle 5*: F-Verteilung bei einem Konfidenzniveau von 99% 288

Tabelle 6*: z-Transformation nach FISHER 289

Nomogramm 7*: Test des einfachen Stichprobenkorrelationskoeffizienten bei einem Konfidenzniveau von 95% ... 290

Nomogramm 8*: Test des multiplen Stichprobenkorrelationskoeffizienten bei einem Konfidenzniveau von 95% ... 291

VERZEICHNIS DER BENUTZTEN SYMBOLE 292

LITERATURVERZEICHNIS 296

LEBENSLAUF 313

VERZEICHNIS DER TABELLEN IM TEXT

4. Kapitel — Seite

Tabelle 4.1 Tabelle der beobachteten Werte für die abhängige Variable y (Verbrauchsmengen) und die unabhängigen Variablen x_1, x_2 und x_3 (Produktionsmengen) für die Beobachtungsperioden $i = 1, 2, \ldots, n$. 39

Tabelle 4.2 Verbrauchsmengen y_i und Produktionsmengen x_{1i}, x_{2i} und x_{3i} von drei Erzeugnissen für 20 Arbeitstage im Beobachtungsmonat. 47

Tabelle 4.3 Differenzen zwischen beobachteten Werten y_i und berechneten Werten y_i' und y_i^+ für die abhängige Variable y. 56

5. Kapitel

Tabelle 5.1 Gesamtkosten y_i und Produktionsmengen x_{1i}, x_{2i} und x_{3i} von drei Erzeugnissen für 10 Beobachtungsperioden. . . 127

Tabelle 5.2 Grenzkosten und ihre Standardfehler für drei Kostenarten zur Ermittlung der optimalen Alternative zwischen Selbstherstellung und Fremdbezug. . 133

6. Kapitel

		Seite
Tabelle 6.1	Zahlenwerte für Kosten (y_i) und Beschäftigung (x_i) für 52 Beobachtungsperioden, geordnet nach der Höhe der Beschäftigung.	170
Tabelle 6.2	Ermittlung der erwarteten Gesamtkosten im Zusammenhang mit der Festlegung eines kritischen Wertes für α (Signifikanzniveau).	185
Tabelle 6.3	Ermittlung der erwarteten Gesamtverluste im Zusammenhang mit der Festlegung eines kritischen Wertes für α (Signifikanzniveau).	185
Tabelle 6.4	Ermittlung der den Kosten K = 100 indifferenten Verluste V bei unterschiedlichem Signifikanzniveau α. .	187
Tabelle 6.5	Selbstkosten je Stück (y_i) und produzierte Stückzahl (x_i) in 15 Betrieben.	193
Tabelle 6.6	Produktionszahlen (x_i) und zugehörige Regreßwerte (y'_i) bei Verwendung einer Parabel der Form $y' = a - bx + cx^2$ als Regressionsfunktion.	196
Tabelle 6.7	Produktionszahlen (x_i) und zugehörige Regreßwerte (y'_i) bei Verwendung einer Hyperbel der Form $y' = a + b\frac{1}{x}$ als Regressionsfunktion.	198
Tabelle 6.8	Tabelle zur Varianzanalyse.	207

VERZEICHNIS DER ABBILDUNGEN IM TEXT

2. Kapitel

Seite

Abb. 2.1 Streuungsdiagramm aus 20 täglichen Beobachtungswerten für die abhängige Variable y (Kosten) bei unterschiedlichen Werten der unabhängigen Variablen x (Produktionsmengen). 14

Abb. 2.2 Darstellung einer Regressionsgeraden und ihrer Parameter im zweidimensionalen Koordinatensystem sowie Angabe der Beobachtungswerte x_i und y_i, des Schätzwertes y'_i und der Abweichungen e_i. 16

Abb. 2.3 Kurvenverläufe von drei in der Kostenrechnung wichtigen Funktionstypen von Regressionslinien. 21

Abb. 2.4 Darstellung einer linearen Regression mit zwei unabhängigen Variablen im dreidimensionalen Koordinatensystem. . . 26

Abb. 2.5 Geometrische Interpretation von Regressionskonstanter und Regressionskoeffizienten einer linearen Regression mit zwei unabhängigen Variablen. 26

4. Kapitel

Abb. 4.1 Erklärung negativer Werte der Regressionskonstanten a bei linearem Verlauf der Regressionslinie im Beobachtungsbereich und nichtlinearem Verlauf in der Nähe der y-Achse. 53

Seite

Abb. 4.2 Approximation einer Kurve durch eine Gerade. 54

Abb. 4.3 Streupunktdiagramm für Beobachtungswerte aus Gemeinkostenbeträgen (y) und Schlüsselmengen der Schlüsselgrößen 1 (x_1) und 2(x_2) für 20 Beobachtungsperioden. . . . 65

Abb. 4.4 Streupunktballung bei 20 Beobachtungswerten für die abhängige Variable y (Kosten) und die unabhängige Variable x (Beschäftigung). 75

5. Kapitel

Abb. 5.1 Streupunktdiagramm aus 20 Beobachtungswerten für die Gemeinkosten (y) und die Beschäftigung (x) zur Ermittlung der Sollgemeinkostenkurve. 85

Abb. 5.2 Verlauf der Gemeinkostenkurve bei Normal- und bei Überbeschäftigung, wenn eine intensitätsmäßige Anpassung an Veränderungen der Beschäftigungslage erfolgt. 90

Abb. 5.3 Statistische Kostenauflösung über die Bestimmung der Regressionsparameter. . . 121

Abb. 5.4 Vertrauensgrenzen einer Regressionsgeraden, gebildet anhand des Standardfehlers der Schätzung $\hat{\sigma}_{y.x}$ und des Standardfehlers der Regressionsgeraden $\hat{\sigma}_{y'}$. 158

6. Kapitel

Seite

Abb. 6.1 Regressionsgerade einschließlich Konfidenzintervall im Abstand einer Standardabweichung und Glockenkurve der Normalverteilung. 173

Abb. 6.2 Darstellung einer Kostenkontrollkarte anhand der Normalverteilung. 176

Abb. 6.3 Kontrollgrenzen bei verschiedenen Werten für α (Signifikanzniveau). 183

Abb. 6.4 Indifferenzkurven für Kosten und Verluste V bei verschiedenen Werten von α . . 188

Abb. 6.5 Stückkosten (y) in Abhängigkeit von der Produktionsmenge (x) für 15 Betriebe. . 194

Abb. 6.6 Regressionsfunktionen für die Abhängigkeit der Stückkosten (y) vom Produktionsvolumen (x). 195

Abb. 6.7 Möglicher Verlauf der Residuen bei Anwendung einer linearen Regressionsfunktion auf einen nichtlinearen (parabolischen) Zusammenhang. 204

Abb. 6.8 Aufspaltung eines beobachteten Stichprobenwertes y_i als Vorbereitung für eine Varianzanalyse. 205

9. Kapitel

Abb. 9.1 Rechenzeiten zur Ermittlung der angegebenen Größen bei Einsatz einer Tischrechenmaschine. 257

1. Kapitel

EINFÜHRUNG

1.1 Problemstellung

Im Rechnungswesen von Industriebetrieben, in dem sich vielschichtige betriebliche Vorgänge zahlenmäßig niederschlagen und durch das die Unternehmensleitung in die Lage versetzt werden soll, das Betriebsgeschehen zu kontrollieren und Anhaltspunkte für Entscheidungen zu gewinnen,[1] haben die Methoden der mathematischen Statistik in den letzten Jahren ständig an Bedeutung gewonnen. Sie ermöglichen das Aufzeigen von Abhängigkeiten und Gesetzmäßigkeiten, deren Kenntnis die Kostenrechnung zu einem wichtigen Instrument der Betriebs- und Unternehmungssteuerung werden läßt.[2] Unter den statistischen Verfahren, die zur Lösung dieser Probleme beitragen, nimmt die Regressions- und Korrelationsanalyse eine bedeutende Stellung ein. Sie ist ein in der Statistik seit langem gebräuchliches Verfahren,[3] um die Abhängigkeit von Größen untereinander zu messen und in mathematischer Form wiederzugeben.

Viele wissenschaftliche Disziplinen, insbesondere die Naturwissenschaften, haben sich relativ früh dieser Methoden bedient. Auch in der Volkswirtschaftslehre wurden zahlreiche Versuche unternommen, um Angebots-, Nachfrage- oder

---

1) K.F. BUSSMANN: Industrielles Rechnungswesen. Stuttgart, 1963. S. 174 f.

2) E. HEINEN: Betriebswirtschaftliche Kostenlehre. Kostentheorie und Kostenentscheidungen. 3. verb. Aufl. Wiesbaden, 1970. S. 36. F. HENZEL: Die Kostenrechnung. 4. erw. u. veränd. Aufl. Essen, 1964. S. 15.

3) Entwickelt wurde die Methode der Regressions- und Korrelationsanalyse ursprünglich im Zusammenhang mit Untersuchungen der Vererbungslehre. Die theoretischen Grundlagen wurden ausführlich von G. Udny YULE im Jahre 1922 dargestellt. Vgl. hierzu: M. EZEKIEL u. K.A. FOX: Methods of Correlation and Regression Analysis. Linear and Curvilinear. Third edition, fifth printing. New York/London/Sydney, 1966. S. 434 f.

Kostenfunktionen empirisch mit Hilfe der Regressionsanalyse abzuleiten.[1)]

Demgegenüber hat man in der Betriebswirtschaftslehre bisher in relativ geringem Maße versucht, diese statistischen Verfahren zu nutzen, sieht man einmal von den empirischen Kostenuntersuchungen in den vierziger Jahren ab, die vornehmlich von J. DEAN, M. EZEKIEL und K.H. WYLIE durchgeführt wurden und die sich die Konstruktion einer gesamtbetrieblichen Kostenfunktion zum Ziel gesetzt hatten.[2)]

In den einzelnen betrieblichen Bereichen dagegen - zum Beispiel im Rechnungswesen und damit auch in der Kostenrechnung - hat die Regressions- und Korrelationsanalyse kaum Beachtung gefunden. Dafür mögen folgende Gründe ausschlaggebend sein:

Die Beobachtung und Aufzeichnung des Datenmaterials ist in der Ökonomie weitaus schwieriger als zum Beispiel in den Naturwissenschaften, wo man sich in einem Experiment die für die Analyse benötigten Daten in der gewünschten Form, d.h. weitgehend frei von störenden Einflüssen anderer als der zu untersuchenden Größen, beschaffen kann.[3)] Die meisten ökonomischen Daten liegen in Form von Zeitreihen vor, d.h. die Beobachtungen müssen über einen Zeitraum hinweg gesammelt werden. Hierdurch ergeben sich für die Regressionsanalyse besondere Probleme, die häufig die Aussagefähigkeit der Ergebnisse beträchtlich einschränken.[4)]

---

1) Vgl. z.B. M. EZEKIEL u. K.A. FOX, a.a.O., S. 444 ff.; Z. HELLWIG: Linear Regression and its Application to Economics. Translated from the Polish by J. STADLER. Oxford/London/New York/Paris und Warszawa, 1963. S. 55 ff.

2) K. WEBER: Die Aussagefähigkeit empirischer Kostenfunktionen in betriebswirtschaftlicher Sicht. "Zeitschrift für Betriebswirtschaft", 37 (1967) 1, 1. Erg. Heft. S. 47 ff. und die dort angegebene Literatur.

3) E. FÖRSTER u. F. EGERMAYER: Korrelations- und Regressionsanalyse. Ein Leitfaden für Ökonomen. Berlin (Ost), 1966. S. 33 f.

4) Vgl. hierzu den Abschnitt 7.1, S. 215 ff.

Ein wichtiger Grund für die relativ seltene Anwendung der Regressions- und Korrelationsanalyse vor allem in der Praxis war bisher sicherlich der große Rechenaufwand, der mit diesen Methoden verbunden ist, insbesondere bei multiplen Analysen, d.h. Untersuchungen mit mehreren unabhängigen Veränderlichen. Wollte man den Rechenaufwand reduzieren, mußte man in der Regel so starke Vereinfachungen des Problems in Kauf nehmen, daß die Ergebnisse keine praktische Bedeutung mehr besaßen.
Schließlich mag auch eine gewisse Gleichgültigkeit und Zurückhaltung gegenüber "komplizierten" statistischen Verfahren - als solches wird die Regressions- und Korrelationsanalyse vielfach angesehen - zu der geringen Anwendung in der betrieblichen Praxis geführt haben.

Erst in den letzten Jahren sind - vornehmlich von amerikanischen Wirtschaftsprüfern und in Deutschland in der Stahlindustrie - Versuche unternommen worden, einige Spezialprobleme der Kostenrechnung mit Hilfe der Regressionsanalyse zu lösen. Auf diese, zum größten Teil in der von der "American Accounting Association" herausgegebenen Zeitschrift "The Accounting Review" und den Zeitschriften der Eisen- und Stahlindustrie veröffentlichten Artikel wird im Laufe dieser Arbeit noch einzugehen sein.

In den deutschsprachigen Standardwerken der Kostenrechnung dagegen findet man keine ausführlichen Darstellungen der Regressions- und Korrelationsanalyse im Zusammenhang mit kostenrechnerischen Problemen. Überwiegend beschränkt man sich auf vereinzelte Hinweise bezüglich der Nützlichkeit dieses statistischen Verfahrens, ohne jedoch näher auf die vielfältigen Grundlagen und Voraussetzungen seiner Anwend-

barkeit einzugehen.[1)]

Es fehlt also immer noch eine systematische Darstellung der Anwendung der Regressions- und Korrelationsanalyse auf kostenrechnerische Probleme, die einen umfassenden Überblick über die Bedeutung dieses Verfahrens gestattet. Daher ist es das Ziel dieser Arbeit, Anwendungsmöglichkeiten der Regressions- und Korrelationsanalyse in der Kostenrechnung der Industrie aufzuzeigen, um Anregungen für die Lösung spezieller, aus der Verschiedenheit der Branchen und Abrechnungsformen resultierender Probleme zu geben.

Naturgemäß können im Rahmen dieser Arbeit nicht alle sich in der Kostenrechnung anbietenden Anwendungsmöglichkeiten ausführlich behandelt werden. Eine solche Aufzählung kann auch nicht der Sinn einer derartigen Arbeit sein. Vielmehr liegt der Schwerpunkt in der Analyse der Voraussetzungen, die für den sinnvollen Einsatz der Regressions- und Korrelationsanalyse in der Kostenrechnung gegeben sein müssen, in der Darstellung der auftretenden Schwierigkeiten und Fehlerquellen und im Aufzeigen der Grenzen des beschriebenen Verfahrens.

Es wird schon an dieser Stelle betont, daß die hier gezeigten statistischen Methoden kein "Allheilmittel" sind, um etwaige kostenmäßige Abhängigkeiten auf einfache Weise in eine mathematische Formel zu kleiden. Sie können aber unter bestimmten Voraussetzungen dann wertvolle Hilfe leisten, wenn im Rahmen

---

1) Vgl. z.B.: K. KÄFER: Standardkostenrechnung. Lehr- und Handbuch. 2. neubearb. u. stark erw. Aufl., Stuttgart, 1964. S. 83, 220, 255, 360.; H. FUNKE, K. MELLEROWICZ u. H.G. ABROMEIT: Grundfragen und Technik der Betriebsabrechnung. 3. Aufl. Freiburg i.Br., 1962. S. 86 f.; W. KILGER: Flexible Plankostenrechnung. Theorie und Praxis der Grenzplankostenrechnung und Deckungsbeitragsrechnung. (Bd. 31 der Veröffentlichungen der SCHMALENBACH-Gesellschaft). 4. Aufl. Köln und Opladen, 1970. S. 366 ff.; A. SCHNETTLER: Betriebsanalyse. Stuttgart, 1958. S. 63 ff.

der Kostenrechnung Kosten zu verrechnen, zu planen oder zu kontrollieren sind. Besonders gewinnbringend können diese Verfahren dort angewendet werden, wo bisher - zum Beispiel im Rahmen der Kostenträgerrechnung bei der Verrechnung der Gemeinkosten auf die Kostenträger - lediglich globale, mehr oder weniger auf dem subjektiven Urteil des Kostenrechners beruhende Verfahren zur Anwendung gelangten.

1.2 Zielsetzung

Bei allen beispielhaften Untersuchungen der folgenden Kapitel wird auf die Bedeutung der Regressions- und Korrelationsanalyse für einzelne Teilgebiete der Kostenrechnung näher eingegangen. Jedoch ist es für das Verständnis dieser Erörterungen sinnvoll, die grundlegende Zielsetzung der Arbeit zunächst zu konkretisieren.

Untersucht man die Anwendungsmöglichkeiten eines mathematisch-statistischen Verfahrens auf einem Gebiet des industriellen Rechnungswesens, so kann eine Beurteilung dieses Verfahrens nur mit Hilfe eines Maßstabs erfolgen, der Aussagen über die Eignung in einem konkreten Fall zuläßt. Zum Beispiel MOEWS hat deutlich gemacht, daß die Kostenrechnung nicht Selbstzweck ist, sondern nur ein Mittel zur Erreichung anderer Zwecke darstellt.[1)] Viel weniger kann also ein statistisches Verfahren wie die Regressions- und Korrelationsanalyse einen eigenen Zweck beinhalten. Sie ist vielmehr ein mathematisches Hilfsmittel, das zur Erreichung bestimmter Zwecksetzungen verwendet werden kann. Bei der vorliegenden Untersuchung bilden die Ziele oder

1) Vgl. D. MOEWS: Zur Aussagefähigkeit neuerer Kostenrechnungsverfahren. (Band 39 der Betriebswirtschaftlichen Forschungsergebnisse). Berlin 1969. S. 35.

Aufgabengebiete der Kostenrechnung den Wertmesser, an dem die Eignung der Regressions- und Korrelationsanalyse in diesem Bereich überprüft werden kann.

Wir wollen zunächst die üblicherweise mit der Kostenrechnung verfolgten Ziele präzisieren. Zwar werden in der Literatur recht unterschiedliche Aufgabengebiete genannt, wobei häufig in Haupt-, Neben- und Sonderzwecke gegliedert wird und unterschiedliche Rangordnungen dieser Zwecke aufgestellt werden,[1] dennoch lassen sich trotz abweichender Systematik und Akzentuierung folgende Ziele als allgemein anerkannte Hauptziele der Kostenrechnung konkretisieren:

1. Ermittlung eines Perioden- oder Stückerfolgs.
2. Planung und Vorgabe von Kosten, wobei sich "Alternativkalküle" (z.B. Wirtschaftlichkeitsrechnungen) und "Zielkalküle" (z.B. Aufstellen von Budgets und Standards) unterscheiden lassen.[2]
3. Kontrolle der Kosten, insbesondere die Kontrolle der betrieblichen Produktivität im innerbetrieblichen (Kostenstellenrechnung) und zwischenbetrieblichen Kostenvergleich.[3]

---

1) Vgl. z.B. H. KOCH: Grundprobleme der Kostenrechnung. Köln und Opladen 1966. S. 37 ff.; D. SCHNEIDER: Die wirtschaftliche Nutzungsdauer von Anlagegütern als Bestimmungsgrund der Abschreibungen. (Band 14 der Beiträge zur betriebswirtschaftlichen Forschung). Köln und Opladen 1961. S. 4 ff.; D. MOEWS, a.a.O., S. 34 ff.; H. MEFFERT: Betriebswirtschaftliche Kosteninformationen. Ein Beitrag zur Theorie der Kostenrechnung. (Band 4 der Schriftenreihe: Die Betriebswirtschaft in Forschung und Praxis. Hrsg. E. HEINEN). Wiesbaden 1968. S. 141 ff.

2) Die Terminologie wurde übernommen von D. SCHNEIDER, Die wirtschaftliche Nutzungsdauer ..., a.a.O. S. 4 ff.

3) Das Aufgabengebiet "Kontrolle" wird vom Verfasser bewußt eng gefaßt. Andere Autoren, z.B. D. SCHNEIDER, rechnen auch die Perioden- und Stückerfolgsrechnung zur Kostenkontrolle. Vgl. D. SCHNEIDER, Die wirtschaftliche Nutzungsdauer ..., a.a.O., S. 10.

Die Frage, inwieweit dieses System von Kostenrechnungszielen konsistent ist, kann hier offengelassen werden, da ihre Beantwortung für die Beurteilung der Regressions- und Korrelationsanalyse im Rahmen der Kostenrechnung ohne Bedeutung ist.[1)]

Dagegen erscheint es dem Verfasser notwendig, den Inhalt der genannten Aufgabengebiete näher zu bestimmen. Allein die Vorgabe dieser Ziele läßt eine gründliche Abwägung der Vor- und Nachteile der Regressions- und Korrelationsanalyse bei ihrer Anwendung auf kostenrechnerische Probleme nicht zu, da alle Ziele relativ abstrakt gefaßt sind. Diese Tatsache erfordert die Aufstellung einer Skala von Beurteilungskriterien, die die genannte Zielsetzung ausfüllen können.

Diese Kriterien ergeben sich aus den Anforderungen, die der Kostenrechner an ein Verfahren zur Berechnung, Verrechnung, Planung oder Kontrolle von Kosten stellen muß. Darum soll im Verlauf dieser Arbeit geprüft werden, inwieweit die Regressions- und Korrelationsanalyse im Rahmen der genannten Aufgabengebiete die Erfüllung folgender Kriterien ermöglicht:

1. O b j e k t i v i t ä t, d.h. die Einschränkung von Subjektivismen und Ermessensspielräumen in der Kostenrechnung.
2. F l e x i b i l i t ä t, d.h. die Möglichkeit zur Anwendung des Verfahrens in den untersuchten Konstellationen ähnlichen Betriebssituationen oder Folgeperioden sowie die Fähigkeit, alternativen Fragestellungen Rechnung zu tragen.

---

1) Vgl. Kapitel 8, in dem der Beweis für diese Behauptung angetreten wird.

3. W i r t s c h a f t l i c h k e i t, d.h. die Überprüfung der Kosten der Durchführung einer solchen Analyse und der Bereitstellung des erforderlichen Datenmaterials.

4. Z u v e r l ä s s i g k e i t, d.h. die Frage nach rechnerischer Richtigkeit der Ergebnisse und nach der Quantifizierung vorhandener Streuungsbereiche dieser Ergebnisse.

5. G e n a u i g k e i t, d.h. insbesondere die Frage nach der Ausdehnung von Bandbreiten oder Streuungsintervallen der aus einem Stichprobenmodell gewonnenen Regressionsergebnisse.

Die in diesem Abschnitt konkretisierten Ziele sollen den Rahmen für die Untersuchungen der Kapitel 4 bis 7 als dem Schwerpunkt der Arbeit darstellen, während die genannten Kriterien bei der folgenden deskriptiven Behandlung der Verfahren den Beurteilungsmaßstab für die Eignung der Regressions- und Korrelationsanalyse in der Kostenrechnung bilden. Im Kapitel 8 schließlich erfolgt die abschließende systematische Beurteilung der Untersuchungsergebnisse im Hinblick auf die hier formulierte Zielsetzung.

2. Kapitel

GRUNDLAGEN DER REGRESSIONS- UND KORRELATIONSANALYSE

2.1 Begriff und Aufgaben

Allgemein versteht man unter dem Begriff "Regressions- und Korrelationsanalyse" mathematisch-statistische Verfahren, die dazu dienen, Abhängigkeiten zwischen zwei oder mehreren quantitativen Merkmalen in mathematischer Form wiederzugeben. Dabei unterscheidet man wegen ihrer verschiedenen Ausgangspunkte und Aufgabenstellungen zwischen der Korrelationsanalyse und der Regressionsanalyse.

Von korrelativen oder stochastischen[1] Zusammenhängen zwischen Erscheinungen spricht man, "wenn bei gegebenen Werten der unabhängigen Veränderlichen sich die Werte der abhängigen Veränderlichen nicht in eindeutiger Weise ergeben, sondern zufallsbedingt in einem Intervall streuen",[2] d.h. sie können mit bestimmten Wahrscheinlichkeiten verschiedene Werte annehmen und bilden keine mathematisch streng definierte Beziehung.[3] Anders formuliert, zu jedem Wert der unabhängigen Veränderlichen gehört eine bestimmte Häufigkeitsverteilung der Werte der abhängigen Veränderlichen.[4] Die Untersuchung dieser korrelativen Zusammenhänge wird als Korrelationsanalyse bezeichnet.

Entgegen diesem wechselseitigen stochastischen Zusammenhang wird die R e g r e s s i o n durch eine e i n s e i t i g e stochastische Abhängigkeit einer zufälligen Veränderlichen von einer oder mehreren anderen Veränderlichen cha-

---

1) Stochastisch im Sinn von "wahrscheinlichkeitstheoretisch"; vgl. auch A.A. TSCHUPROW: Principles of the Mathematical Theory of Correlation. Translated by M. KANTOROWITSCH. London/Edinburgh und Glasgow, 1939. S. 25 Fußnote.

2) E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 15.

3) H. NIKLAS u. M. MILLER: Korrelationsrechnung und ihre Anwendung auf Statistik, Versuchswesen, Vererbungslehre, Wissenschaft und Technik. Leipzig (1940). S. 1.

4) E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 15 f.

rakterisiert,[1] die nicht zufällige Veränderliche zu sein brauchen, sondern deren Werte vor Durchführung der Analyse festgelegt werden können.[2]

Häufig existieren Beziehungen zwischen Veränderlichen, zwischen denen schon auf Grund logischer Überlegungen nur eine solche einseitige Abhängigkeit sinnvoll ist. So kann man beispielsweise die Kosten eines Betriebes in Abhängigkeit von der Faktorqualität, den Faktorpreisen und den Faktorproportionen - um nur einige Kosteneinflußgrößen zu nennen - darstellen, nicht aber von einer wechselseitigen Abhängigkeit sprechen wie zum Beispiel bei einem Zusammenhang zwischen Körpergewicht und Körpergröße von Personen. Untersucht man die Kosten und ihre Einflußgrößen, so sind schon von der Aufgabenstellung her eindeutig die abhängige und die unabhängige(n) Variable(n) festgelegt, so daß auch ein Problem der Umkehrbarkeit der Regression nicht auftreten kann.[3] Die Untersuchung dieser einseitigen stochastischen Abhängigkeit bezeichnet man als Regressionsanalyse.

Allerdings bestehen zwischen Korrelation und Regression enge Beziehungen, so daß man - insbesondere in älteren Veröffentlichungen - häufig die Regression als Sonderfall der Korrelation angesehen und den Begriff Korrelation als Oberbegriff verwendet hat.[4] Heute werden die Regressions- und die Korrelationsrechnung in der Regel gleichberechtigt nebeneinander gestellt,[5] teilweise zeichnet sich sogar eine Schwerpunkt-

---

1) E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 25.; A. HALD: Statistical Theory with Engineering Applications. Fifth printing. New York/London, 1962. S. 613 f.

2) M. EZEKIEL u. K.A. FOX, a.a.O., S. 130.; F. EGERMAYER: Die Korrelationsanalyse in der ökonomischen Forschung. "Zeitschrift für Messen, Steuern, Regeln". O.J. (1962) S. 435 ff., hier S. 437.

3) E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 26 f.

4) Ebenda. S. 27.

5) Vgl. z.B. die Titel der älteren und neueren Ausgaben des Werkes von M. EZEKIEL bzw. M. EZEKIEL u. K.A. FOX; M. EZEKIEL: Methods of Correlation Analysis. Second edition. New York, 1953.; M. EZEKIEL u. K.A. FOX: Methods of Correlation and Regression Analysis. Third edition. New York/London/Sydney, 1966.

verlagerung auf die Regression hin ab, da sie noch in Fällen Anwendung finden kann, in denen die Voraussetzungen für die Durchführung einer Korrelationsanalyse nicht mehr gegeben sind.[1)]

Die Aufgaben, die im allgemeinen an die Regressions- und Korrelationsanalyse gestellt werden, lassen sich folgendermaßen gliedern:[2)]

1. Aufgaben der Korrelationsanalyse
   a. Messung der E n g e oder "S t r a m m h e i t" des Zusammenhangs zwischen den beobachteten Erscheinungen.
   b. Messung des Grades des korrelativen Zusammenhangs zwischen mehreren Erscheinungen, um aus der normalerweise vorhandenen Vielzahl von unabhängigen Variablen die wichtigsten, d.h. die mit dem größten Einfluß auf die Zielgröße, herauszusuchen. Das sind jene Faktoren, die mit der Zielgröße den engsten Zusammenhang bilden.

2. Aufgaben der Regressionsanalyse
   a. Ermittlung der A r t oder F o r m des untersuchten Zusammenhangs.
   b. Schätzung noch unbekannter Werte der Zielgröße auf Grund von verschiedenen möglichen Werten der Einflußgrößen, einmal als zukünftige Werte für Prognosezwecke, zum anderen als "Zwischenwerte" für Verteilungs- oder Zurechnungszwecke, wenn es sich um mehrere unabhängige Variable handelt.

---

1) Es ist nur eine einseitige stochastische Abhängigkeit zur Durchführung der Regressionsanalyse erforderlich. Vgl. S. 9 f.

2) In Anlehnung an E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 29 ff. u. S. 103.; E. FÖRSTER: Zur Anwendung der Korrelations- und Regressionsanalyse in der Ökonomie. "Wissenschaftliche Zeitschrift der Humboldt-Universität zu Berlin". Gesellschafts- und sprachwissenschaftliche Reihe. 13 (1964) 5, S. 559 ff., hier S. 560 f.; O. ANDERSON: Probleme der statistischen Methodenlehre in den Sozialwissenschaften. 5. unveränd. Aufl. Würzburg 1965. S.186 f.

Aus dieser kurzen Charakterisierung der Aufgaben der Korrelationsanalyse und der Regressionsanalyse geht bereits hervor, daß für die im Rahmen dieser Arbeit durchgeführten Kostenuntersuchungen der Regressionsanalyse - und hier insbesondere ihrer multiplen Form - ein größeres Gewicht zuzumessen ist; denn häufig will man einen mathematischen Ausdruck für eine Kostenabhängigkeit finden oder Kosten verteilen und zurechnen. Hier bietet eine Regressionsfunktion wertvolle Anhaltspunkte. Daneben wird aber auch gezeigt werden, auf welche Weise die Korrelationsrechnung über die Bestimmung von Korrelationskoeffizienten oder Bestimmtheitsmaßen zur Lösung von kostenrechnerischen Problemen beitragen kann.

Schon an dieser Stelle sei kurz auf das Problem "Korrelation und Kausalität" eingegangen. Es muß nachdrücklich betont werden, daß der statistische Begriff der Korrelation im Hinblick auf die Frage der Kausalität völlig neutral ist.[1] Ein hoher Korrelationskoeffizient sagt lediglich aus, daß die Variation in der einen Variablen in hohem Maße mit der Variation in der anderen Variablen v e r b u n d e n ist.[2] Ob sie jedoch durch die Variation der anderen Veränderlichen v e r u r s a c h t wurde, ist eine Frage, die nicht durch die Korrelationsanalyse selbst, sondern nur durch außerhalb der statistischen Rechnung liegende Untersuchungen oder Über-

---

1) R.G.D. ALLEN: Statistik für Volkswirte. Aus dem Englischen übertragen von W. FÖRSTER. 2. unveränd. Aufl. Tübingen, 1964. S. 125.

2) M. EZEKIEL und K.A. FOX, a.a.O., S. 139 f.; W.A. WALLIS und H.V. ROBERTS: Methoden der Statistik. Ein neuer Weg zu ihrem Verständnis. Übersetzt von H. von WALDHEIM. Mit einem Vorwort von R. WAGENFÜHR. Freiburg i. Br. (1960). S. 463.

legungen beantwortet werden kann.[1] Aus diesem Grund ist eine Kausalitätsaussage allein auf der Basis der durch die Korrelationsanalyse ermittelten Zahlen problematisch und birgt die Gefahr der Interpretation sogenannter "Nonsens-Korrelationen" in sich.

2.2 Schätzungen nach der Methode der kleinsten Quadrate

In der Literatur werden verschiedene Verfahren beschrieben, die es ermöglichen, aus den Werten von unabhängigen Variablen unbekannte Werte der Zielgröße zu schätzen, u.a. die Methode des günstigsten Anfangspunktes,[2] die Rangkorrelation,[3] die Zwei-Punkte-Methode[4] oder graphische Methoden.[5]

In der vorliegenden Arbeit wird jedoch ausschließlich das Schätzverfahren nach der Methode der kleinsten Quadrate angewendet, da es das gebräuchlichste ist, bei linearen Funktionen konsistente, unverzerrte und wirksame Schätzwerte liefert[6] und auch Regressions- und Korrelationsprogramme für elektronische Datenverarbeitungsanlagen fast ausschließlich auf diesem Prinzip aufgebaut sind.[7]

---

1) A. HALD, a.a.O., S. 614.; R.G.D. ALLEN, a.a.O., S. 125.; E.W. GAYNOR: Statistical Testing of Relationships Used in Cost Planning and Allocation. "National Association of Cost Accountants (N.A.C.A.) - Bulletin". Section 1. New York, Mai 1957, S. 1114 ff., hier S. 1115 f.

2) Vgl. hierzu: E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 67 ff.

3) Vgl. z.B. M.G. KENDALL: The Advanced Theory of Statistics. Vol. 1. Third edition. London 1947. S. 388 ff.

4) Vgl. z.B. Z. HELLWIG, a.a.O., S. 124 ff.

5) Vgl. z.B. M. EZEKIEL u. K.A. FOX, a.a.O., S. 210 ff. u. S. 254 ff.; E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 145 ff.

6) Z. HELLWIG, a.a.O., S. 98.; T. YAMANE: Statistics, an Introductory Analysis. New York, Evanston, London u. Tokyo 1964. S. 377.

7) Vgl. 9. Anhang I: Rechnerische Durchführung von Regressions- und Korrelationsanalysen S. 252 ff., insbesondere den Gliederungspunkt 9.3, S. 260 ff.

Die Grundlagen der Methode der kleinsten Quadrate - genauer bezeichnet als Methode, bei der die Summe der Abweichungsquadrate minimiert wird - sollen im folgenden dargestellt werden.

Wir nehmen an, daß wir während eines bestimmten Zeitraumes die für die abhängige Variable y (z.B. Kosten) und die unabhängige Variable x (z.B. Produktionsmenge) beobachteten Werte aufgezeichnet und in ein Schaubild eingetragen haben. Dadurch erhalten wir ein sogenanntes "Streuungsdiagramm".[1)]

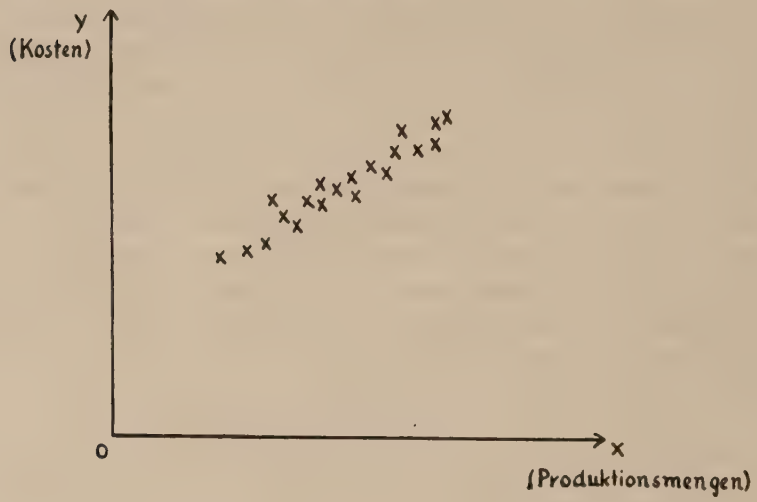


Abb. 2.1: Streuungsdiagramm aus 20 täglichen Beobachtungswerten für die abhängige Variable y (Kosten) bei unterschiedlichen Werten der unabhängigen Variablen x (Produktionsmengen).

1) Auch als "Korrelationsdiagramm" oder "Korrelationsbild" bezeichnet.

Jedes Kreuz in diesem Streuungsdiagramm repräsentiert ein beobachtetes xy-Wertepaar. Unser Ziel ist es, durch diese Punktwolke eine Linie zu legen, die die Beziehung zwischen y und x "bestmöglich" wiedergibt.

Bei der Methode der kleinsten Quadrate wählen wir einen allgemeinen Gleichungsansatz der Form

$$y_i = f(x_i) + e_i, \qquad (2.1)$$

wobei y die abhängige Variable, x die unabhängige Variable und die e_i jene störenden Komponenten der Reihe sind, die sich nicht durch die Regressionsfunktion erklären lassen und unter Umständen als durch den Zufall bedingt angesehen werden können.[1)] Zur Unterscheidung von den beobachteten Werten der abhängigen Variablen - y_i - werden im folgenden die durch die Funktion gelieferten Schätzwerte für die abhängige Variable mit y_i' bezeichnet, d.h. y_i ist die y-Koordinate eines beliebigen Punktes P_i im Streuungsdiagramm, während y_i' immer die y-Koordinate eines Punktes auf der Regressionslinie ist.

Um eine Linie zu finden, die die beobachteten Werte bestmöglich wiedergibt, d.h. die die Abweichungen e_i zwischen den beobachteten Werten y_i und den durch die Schätzfunktion ermittelten Werten y_i' möglichst klein werden läßt, bildet man die Abweichungsquadrate

$$e_i^2 = (y_i - y_i')^2,$$

summiert sie und sucht das Minimum:

$$S = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - y_i')^2 = \sum_{i=1}^{n} \left[y_i - f(x_i) \right]^2. \qquad (2.2)$$

---

1) Vgl. O. ANDERSON, a.a.O., S. 196.; G. GRÖNER: Ein Beitrag zur Anwendung der Mehrfachkorrelation auf wirtschaftliche Reihen. Diss. an der Wirtschaftshochschule Mannheim. Mannheim 1957. S. 10.

Für die einfache lineare Regressionsanalyse mit der Regressionsgleichung

$$y'_i = a_{yx} + b_{yx}x_i \qquad (2.3)$$

lautet die Summe S:

$$S = \sum_{i=1}^{n} (y_i - a_{yx} - b_{yx}x_i)^2 = \text{Min.} \qquad (2.4)$$

Zur Veranschaulichung mag die folgende Darstellung dienen, die auch zeigt, daß im Fall der hier betrachteten yx-Regressionsfunktion die Abweichungen parallel zur y-Achse gebildet werden.[1)]

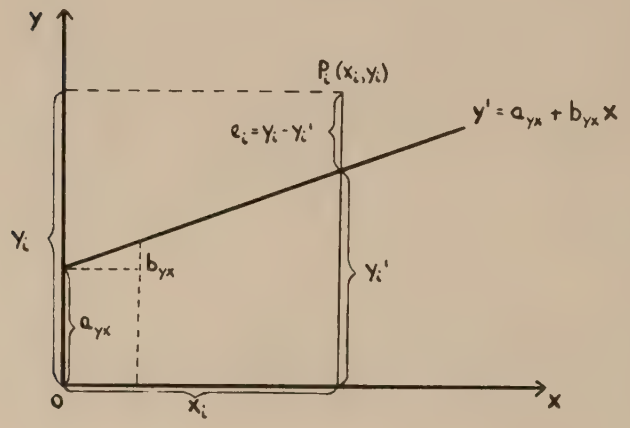


Abb. 2.2: Darstellung einer Regressionsgeraden und ihrer Parameter im zweidimensionalen Koordinatensystem sowie Angabe der Beobachtungswerte x_i und y_i, des Schätzwertes y'_i und der Abweichungen e_i.

1) Will man aus gegebenen y_i-Werten Schätzwerte für die x_i ermitteln, so minimiert man die Abweichungsquadrate parallel zur x-Achse und ermittelt die xy-Regressionsfunktion. Wenn Regressionsfunktionen bestimmt werden, bei denen die Abweichungen der beobachteten Werte von den Regresswerten durch senkrechte Abstände der Punkte von der Regressionslinie gebildet werden, spricht man von "orthogonalen Regressionen". Sie sollen im Rahmen dieser Arbeit ausgeklammert werden.

Mathematisch formuliert stellt sich das Problem folgendermaßen:
Den Ausgangspunkt bildet die Gleichung (2.4), wobei die y_i- und die x_i-Werte durch unsere Aufzeichnungen bekannt sind und die Parameter a_{yx} und b_{yx} errechnet werden sollen. Die Summe der Abweichungsquadrate S ist - wie aus Gleichung (2.4) hervorgeht - eine Funktion der Parameter a_{yx} und b_{yx}, so daß man schreiben kann:

$$S = f\,(a_{yx},\ b_{yx})\ . \qquad (2.5)$$

Die aus der Mathematik bekannten Bedingungen für das Minimum einer solchen Funktion sind dann erfüllt, wenn die ersten partiellen Ableitungen gleich Null sind und die zweiten Ableitungen ein positives Vorzeichen haben.[1)]

Durch partielle Differentiation von Gleichung (2.4) nach a_{yx} und b_{yx} erhält man:

$$\frac{d\,S}{d\,a_{yx}} = -\,2 \sum_{i=1}^{n} (y_i - a_{yx} - b_{yx}x_i) = 0 \quad \text{und} \qquad (2.6)$$

$$\frac{d\,S}{d\,b_{yx}} = -\,2 \sum_{i=1}^{n} (y_i - a_{yx} - b_{yx}x_i)x_i = 0\ . \qquad (2.7)$$

Nach Division der Gleichungen (2.6) und (2.7) durch - 2, Ausmultiplikation und Auflösung der Klammern erhält man die sogenannten "Normalgleichungen", aus denen sich die Werte für a_{yx} und b_{yx} bestimmen lassen:

---

1) Zur Darstellung der Vorgehensweise soll hier die Bildung der ersten partiellen Ableitungen genügen. Der mathematisch interessierte Leser findet eine ausführliche Darstellung bei E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 58.

$$\sum_{i=1}^{n} y_i - n\, a_{yx} - b_{yx} \sum_{i=1}^{n} x_i = 0 \quad \text{und} \qquad (2.8)$$

$$\sum_{i=1}^{n} x_i y_i - a_{yx} \sum_{i=1}^{n} x_i - b_{yx} \sum_{i=1}^{n} x_i^2 = 0 \,. \qquad (2.9)$$

Sie werden häufig auch in der folgenden Form angegeben:[1)]

$$\sum_{i=1}^{n} y_i = n\, a_{yx} + b_{yx} \sum_{i=1}^{n} x_i \quad \text{und} \qquad (2.10)$$

$$\sum_{i=1}^{n} x_i y_i = a_{yx} \sum_{i=1}^{n} x_i + b_{yx} \sum_{i=1}^{n} x_i^2 \,. \qquad (2.11)$$

Die auf diese Weise ermittelte Gerade bezeichnet man als die "Linie der besten Anpassung" (line of best fit), da sie für alle beobachteten Werte von x Schätzwerte für y liefert, die so nahe wie möglich an alle verschiedenen beobachteten y-Werte herankommen.[2)] Denn die Summe der Abweichungsquadrate ist kleiner als sie sein würde, wenn irgendeine andere Gerade verwendet worden wäre.[3)]

Um auf dem beschriebenen Weg zu sinnvollen Ergebnissen zu gelangen,[4)] ist Voraussetzung, daß die Abweichungen e_i - auch als "Fehler" oder "Residuen" bezeichnet - im wahrscheinlichkeitstheoretischen Sinn voneinander unabhängig sind und

---

1) Auf weitere Umformungen oder Auflösungen dieser Normalgleichungen kann verzichtet werden. Bei umfangreichen Analysen empfiehlt sich die Verwendung eines Rechenschemas. Vgl. dazu 9. Anhang I über die rechnerische Durchführung von Regressions- und Korrelationsanalysen, S. 252 ff.

2) M. EZEKIEL u. K.A. FOX, a.a.O., S. 63.

3) Ebenda. Fußnote.

4) Unter "sinnvollen Ergebnissen" versteht der Verfasser die Ermittlung von Schätzwerten, die durch die Bildung von Konfidenzintervallen "statistisch gesichert" werden können.

konstante endliche Streuung haben.[1] Sind diese Voraussetzungen nicht erfüllt, dann treten Probleme von Autokorrelation und Heteroskedasie auf, die im Verlaufe dieser Arbeit noch darzustellen sein werden.[2]

Die mit Hilfe der Methode der kleinsten Quadrate ermittelten Parameter a_{yx} und b_{yx} werden üblicherweise als "Regressionskoeffizienten" bezeichnet. In dieser Arbeit wird im folgenden nur b_{yx} als Regressionskoeffizient bezeichnet, während für das absolute Glied a_{yx} der Begriff "Regressionskonstante" verwendet wird.

Der Regressionskoeffizient b_{yx} gibt geometrisch gesehen die Steigung der Regressionsgeraden wieder und sagt damit aus, wie groß die durchschnittliche Veränderung der abhängigen Variablen ist, wenn der Wert der unabhängigen Veränderlichen um eine Einheit verändert wird. Im Englischen wird er manchmal als "gross regression coefficient" bezeichnet,[3] um damit anzuzeigen, daß er die Relation zwischen y und x wiedergibt, ohne zu berücksichtigen, ob die Veränderungen in y allein auf Veränderungen in x zurückzuführen sind oder auch auf andere unabhängige Variable außer x.[4]

---

1) G. TINTNER: Handbuch der Ökonometrie. (Enzyklopädie der Rechts- und Staatswissenschaft, begr. von F. von LISZT und W. KASKEL, Hrsg. W. Kunkel, H. Peters und E. Preiser. Abteilung Staatswissenschaft). Berlin/Göttingen/Heidelberg, 1960. S. 209.
J. TINBERGEN: Einführung in die Ökonometrie. Mit einem Geleitwort von E. SCHNEIDER. Ins Deutsche übertragen von H. CHRISTOPH. Vom Verf. durchges. u. erg.; Wien und Stuttgart, 1952. S. 99 f.

2) Vgl. S. 215 ff. u. S. 223 ff.

3) M. EZEKIEL u. K.A. FOX, a.a.O., S. 154.

4) Vgl. dazu auch die Interpretation der partiellen Regressionskoeffizienten. S. 24 f.

2.3 Nichtlineare Regressionsfunktionen

Im Abschnitt 2.2 wurde das Schätzverfahren nach der Methode der kleinsten Quadrate für den einfachsten Fall einer l i n e a r e n Einfachregression dargestellt. Jedoch wird bei praktischen Untersuchungen nicht immer die Anwendung einer linearen Funktion sinnvoll sein. Häufig ist auf Grund früherer Untersuchungen oder aus der Theorie bekannt, daß ein linearer Gleichungsansatz nicht zu befriedigenden Ergebnissen führen kann. Auch aus dem Verlauf der Punktwolke in einem Streuungsdiagramm läßt sich oft eine eindeutig nichtlineare Tendenz erkennen, die die Berechnung einer Regressionsgeraden nicht sinnvoll erscheinen läßt.

In diesen Fällen muß versucht werden, eine nichtlineare Gleichung zu finden, durch die die Beobachtungswerte möglichst gut repräsentiert werden. In aller Regel ergeben sich dabei Schwierigkeiten in Bezug auf den zu wählenden Funktionstyp, da e i n e Form aus den vielen Möglichkeiten nichtlinearer Gleichungsansätze auszuwählen ist.

Schon die Darstellung einer ausgewählten Anzahl von Regressionkurven,[1)] denen eine gewisse praktische Bedeutung zukommt, läßt die Schwierigkeiten erkennen, die die Wahl der r i c h t i g e n Kurve mit sich bringt.

Um zu praktisch verwertbaren Ergebnissen zu gelangen, muß sich der Untersuchende stets über die für sein Problem theoretisch möglichen Kurvenverläufe Rechenschaft ablegen, d.h. er muß Überlegungen anstellen, ob der betrachtete Zusammenhang sich durch eine monoton steigende oder fallende Kurve ausdrücken läßt, ob vielleicht an einer Stelle ein Wendepunkt denkbar ist oder ob mit einem Minimum oder einem Maximum im Kurvenverlauf zu rechnen ist.

---

1) Vgl. Abbildung 2.3 auf S. 21.

Da auf diese Problematik in dieser Arbeit noch ausführlich einzugehen sein wird, mag an dieser Stelle der gegebene Überblick vorerst genügen.

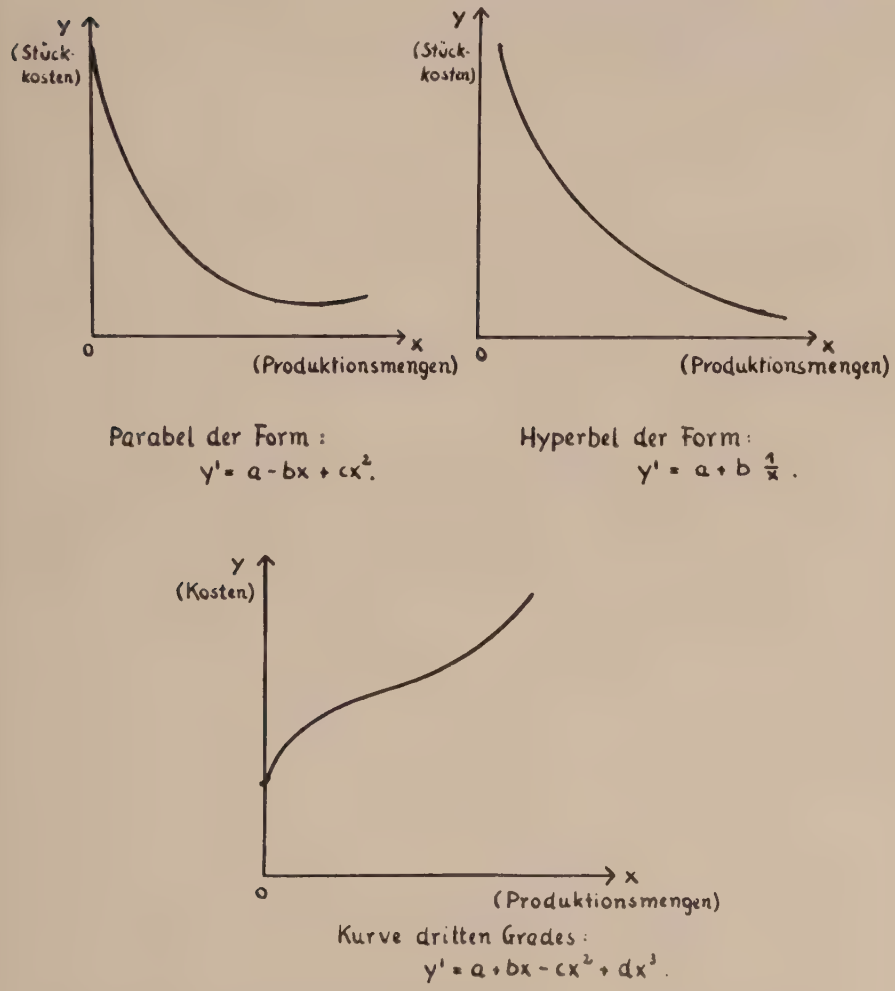


Abb. 2.3: Kurvenverläufe von drei in der Kostenrechnung wichtigen Funktionstypen von Regressionslinien.

Wie bei dem linearen Ansatz der Regressionsgleichung

$$y' = a_{yx} + b_{yx}x$$

kann auch bei nichtlinearem Ansatz die Methode der kleinsten Quadrate Verwendung finden, um die Parameter der Regressionsfunktionen zu bestimmen.

Für eine Parabel der Form

$$y' = a_{yx} + b_{yx}x + c_{yx}x^2 \qquad (2.12)$$

erhält man die Summe der Abweichungsquadrate analog zu dem im Abschnitt 2.2 beschriebenen Verfahren:

$$S = \sum_{i=1}^{n} (y_i - y_i')^2 = \sum_{i=1}^{n} (y_i - a_{yx} - b_{yx}x_i - c_{yx}x_i^2)^2$$

$$= \text{Min.} \qquad (2.13)$$

Durch partielle Differentiation nach den drei Variablen a_{yx}, b_{yx} und c_{yx} und durch verschiedene Umformungen ergeben sich die folgenden drei Normalgleichungen, aus denen sich die Werte für die Parameter der Regressionsgleichung errechnen lassen:[1)]

$$\sum y_i = n\,a_{yx} + b_{yx}\sum x_i + c_{yx}\sum x_i^2, \qquad (2.14)$$

$$\sum x_i y_i = a_{yx}\sum x_i + b_{yx}\sum x_i^2 + c_{yx}\sum x_i^3 \quad \text{und} \qquad (2.15)$$

$$\sum x_i^2 y_i = a_{yx}\sum x_i^2 + b_{yx}\sum x_i^3 + c_{yx}\sum x_i^4 . \qquad (2.16)$$

Für die auf Seite 21 angegebenen Funktionstypen errechnen sich die entsprechenden Normalgleichungen in analoger Weise, so daß auf ihre Ableitung verzichtet werden kann.

---

1) Auf die Angabe der Summationsgrenzen wird hier wegen einer größeren Übersichtlichkeit der Schreibweise verzichtet. Alle Werte von y und x laufen von i = 1, 2, ..., n.

Inwieweit sich gekrümmte Linien für ein begrenztes Kurvenstück durch eine Gerade approximieren lassen oder inwieweit durch logarithmische Transformationen Verschiebungen in Richtung auf die Linearität einer Funktion erreicht werden können, soll nicht an dieser Stelle, sondern anhand von Beispielen in den Hauptteilen der Untersuchung dargestellt werden.[1)]

2.4 Multiple und partielle Regression

In vielen Fällen reicht eine einzige Einflußgröße nicht aus, die Veränderungen in der Zielgröße zu erklären. Die Zusammenhänge zwischen ökonomischen Erscheinungen stellen sich nur selten als einfache, d.h. als Zusammenhänge zwischen einer abhängigen und e i n e r unabhängigen Variablen dar. In aller Regel wirkt eine Vielzahl von einflußnehmenden Faktoren auf die Zielgröße.[2)] Besonders im Rahmen der Wirtschaftswissenschaften sind kaum "Laborsituationen" anzutreffen, in denen man den Einfluß aller Faktoren bis auf den einen zu untersuchenden weitgehend ausschalten kann, so daß eine einfache Analyse nicht anwendbar ist. Es muß vielmehr das Gesamtresultat der wichtigsten in der bestehenden Situation wirkenden Faktoren zur gleichen Zeit gemessen werden.[3)] Es kann nämlich der Fall eintreten, daß bei einer einfachen Analyse eine bestimmte Ursache (unabhängige Variable) keinen sichtbaren Einfluß auf die Zielgröße (abhängige Variable) hat, weil neben der betrachteten Variablen noch eine oder mehrere andere unabhängige Variable einen Einfluß auf die

1) Ausführliche Kurvendiskussionen findet man z.B. bei M. EZEKIEL und K.A. FOX, a.a.O., S. 70 ff.; F.E. CROXTON u. D.J. COWDEN: Applied General Statistics. Second edition. London (1965). S. 282 ff.

2) Vgl. auch E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 80 ff.

3) M. EZEKIEL u. K.A. FOX, a.a.O., S. 152; G. GRÖNER, a.a.O., S. 9.

Zielgröße ausüben, der den Einfluß der zuerst betrachteten Veränderlichen aufhebt.[1] In solchen Fällen würde man bei Durchführung einer einfachen Regressions- und Korrelationsanalyse zu falschen Schlußfolgerungen gelangen.

Im Fall der multiplen Regressionsanalyse wählt man einen Gleichungsansatz mit mehreren unabhängigen Veränderlichen, zum Beispiel bei linearer Analyse mit zwei unabhängigen Variablen:[2]

$$y' = a_{y.12} + b_{y1.2}x_1 + b_{y2.1}x_2 \quad . \qquad (2.17)$$

Diese Schreibweise soll auch in den folgenden Abschnitten beibehalten werden, da sie übersichtlich ist und durch die Stellung der Ziffern in den Subskripten die ausgedrückten Abhängigkeiten sogleich erkennen läßt:

Die Subskripte bei a geben an, "daß es sich bei a um eine Konstante in einer Regressionsfunktion handelt, die die Abhängigkeit der Veränderlichen y von den Veränderlichen x_1 und x_2 darstellt".[3] Der Einfachheit halber werden die Veränderlichen nur durch die Ziffern 1 bzw. 2 repräsentiert. Der Punkt im Subskript der Regressionsparameter soll kennzeichnen, welche Regression bezüglich welcher Abhängigkeit untersucht wird.[4]

Die Koeffizienten $b_{y1.2}$ und $b_{y2.1}$ werden als multiple Regressionskoeffizienten bezeichnet. Da zwischen der multiplen und der partiellen Regression keine methodischen Unterschiede bestehen,[5] nennt man $b_{y1.2}$ und $b_{y2.1}$ auch partielle Regressionskoeffizienten, in der angloamerikanischen

---

1) E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 80.

2) Ebenda. S. 80 ff.; ähnliche Schreibweise bei M. EZEKIEL u. K.A. FOX, a.a.O., S. 176.

3) E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 80 f.

4) Ebenda. S. 81.

5) Im Gegensatz zur multiplen und partiellen K o r r e l a t i o n. Vgl. S. 29 f.

Literatur "net regression coefficients" oder "partial regression coefficients".[1)]

Der Regressionskoeffizient $b_{y1.2}$ ($b_{y2.1}$) gibt an, wie sich die Variable y durchschnittlich verändert, wenn sich die Variable x_1 (x_2) um eine Einheit ändert und die Variable x_2 (x_1) konstant bleibt.[2)]

Um für die Funktion $y' = a_{y.12} + b_{y1.2}x_1 + b_{y2.1}x_2$ die numerischen Werte zu bestimmen, bedient man sich analog zu der Vorgehensweise für die einfache Regressionsanalyse der Methode der kleinsten Quadrate. Da man für jede weitere unabhängige Variable eine weitere Reihe von Beobachtungswerten zur Verfügung hat, erhält man aus der Auflösung der Ausgangsgleichung

$$S = \sum_{i=1}^{n} (y_i - y_i')^2 = \text{Min.}$$

eine weitere Normalgleichung, durch deren Auflösung sich die numerischen Werte der Regressionskonstanten und der Regressionskoeffizienten ergeben.

Wie die Abhängigkeit einer Zielgröße von einer Einflußgröße, so läßt sich auch die Abhängigkeit einer Zielgröße von zwei Einflußgrößen graphisch veranschaulichen. Die Beobachtungswerte bilden in diesem Fall eine Punktwolke im dreidimensionalen Raum, wobei die Regressionsfunktion

$$y' = a_{y.12} + b_{y1.2}x_1 + b_{y2.1}x_2$$

bei linearer Abhängigkeit als Regressionsebene einen Schnitt durch die Punktwolke darstellt.[3)]

---

1) M. EZEKIEL u. K.A. FOX, a.a.O., S. 153 u. S. 176.

2) E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 81.

3) Vgl. auch E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 85 f.

Zeichnet man die Regressionsfunktion (2.17) als yx_1x_2 - Ebene in ein Schaubild ein, so erhält man folgende Darstellung:

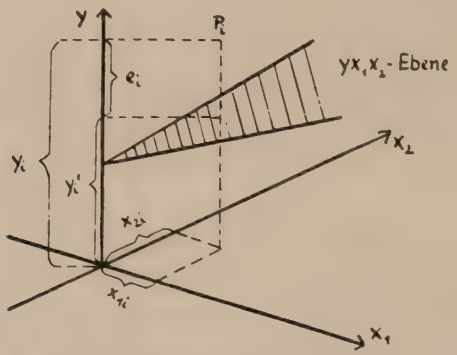


Abb. 2.4: Darstellung einer linearen Regression mit 2 unabhängigen Variablen im dreidimensionalen Koordinatensystem.

Zur geometrischen Interpretation der Regressionskonstanten und der beiden partiellen Regressionskoeffizienten mag schließlich die folgende Darstellung dienen:

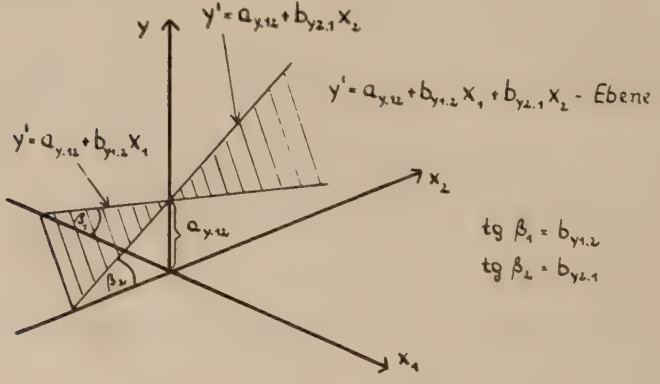


Abb. 2.5: Geometrische Interpretation von Regressionskonstanter und Regressionskoeffizienten einer linearen Regression mit zwei unabhängigen Variablen.

Eine vierdimensionale Regression, d.h. eine Regressionsfunktion mit drei unabhängigen Variablen ist graphisch zwar nicht mehr darstellbar, aber rechnerisch in gleicher Weise wie die beschriebene dreidimensionale Regression zu ermitteln. Erweitert man die Gleichung (2.17) zu einer Regressionsgleichung mit m unabhängigen Veränderlichen, so erhält man die allgemeine Formulierung des Problems für eine (m +1)-dimensionale Regression:[1)]

$$y' = a_{y.12...m} + b_{y1.23...m}x_1 + \dots + b_{yk.12...(k-1)(k+1)...m}x_k + \dots + b_{ym.12...m-1}x_m \quad (2.18)$$

$$\text{für } k = 1, 2, \dots, m.$$

2.5 Der Korrelationskoeffizient

Die Frage der Stärke der Relation der abhängigen Variablen zur unabhängigen Variablen wird mit Hilfe des Korrelationskoeffzienten (bzw. des Bestimmtheitsmaßes) beantwortet. Üblicherweise bezeichnet man den Korrelationskoeffizienten mit r_{yx} - wenn die Abhängigkeit der Variablen y von der Variablen x untersucht wird - und das Bestimmtheitsmaß, das gleich dem Quadrat des Korrelationskoeffizienten ist, mit r^2_{yx} oder mit d_{yx}.[2)] Dieses Maß sagt aus, welcher Anteil der Veränderungen in den Werten der abhängigen Veränderlichen durch die Veränderungen in den Werten der unabhängigen Veränderlichen "erklärt" werden oder auf diese Veränderungen zurückgeführt werden kann.[3)] Der Korre-

1) Vgl. E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 88.

2) Die Bezeichnung des Bestimmtheitsmaßes mit dem Buchstaben "d" stammt von der englischen Bezeichnung dieses Maßes als "coefficient of determination".

3) Vgl. z.B. W.A. WALLIS u. H.V. ROBERTS, a.a.O., S. 463.

lationskoeffizient und das Bestimmtheitsmaß sind dimensionslos und damit unabhängig von der gewählten Einheit, in der die Werte der Zielgröße ausgedrückt werden.[1)]

Der Wert des einfachen Korrelationskoeffizienten kann zwischen -1 und +1 schwanken - die Grenzen mit eingeschlossen -. Im Fall r = -1 spricht man von perfekter negativer, im Fall r = +1 von perfekter positiver Korrelation zwischen den Veränderlichen, im Fall r = 0 von völlig fehlender Korrelation.

Die allgemeine Formel für den Korrelationskoeffizienten lautet:[2)]

$$r_{yx} = \sqrt{\frac{\sum_{i=1}^{n} (y_i' - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}} \quad . \qquad (2.19)$$

Bei verbaler Interpretation der Formel (2.19) ergibt sich, daß der so definierte Korrelationskoeffizient gleich der Quadratwurzel des Quotienten aus dem durch die Regressionsfunktion erklärten Teilbetrag an der Gesamtstreuung der y_i-Werte $(y_i' - \bar{y})^2$ und der Gesamtstreuung der y_i-Werte $(y_i - \bar{y})^2$ ist. Sowohl der erklärte Teilbetrag als auch die Gesamtstreuung werden durch die jeweiligen Summen der quadratischen Abweichungen gemessen. Hieraus erklärt sich auch der häufig verwendete Satz, daß das Bestimmtheitsmaß den Anteil der Veränderungen in den Werten der abhängigen Variablen angibt, der durch die Veränderungen in den Werten der unabhängigen Variablen erklärt wird.[3)]

Auf weitere Umformungen der Gleichung (2.19), so insbesondere für Zwecke der praktischen Berechnung, und auf die

---

1) F.E. CROXTON u. D.J. COWDEN, a.a.O., S. 461.

2) Vgl. E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 110. $\bar{y}$ ist definiert als arithmetisches Mittel aus den Werten der abhängigen Variablen.

3) Vgl. S. 27.

Darstellung des sogenannten "Produktmomentkorrelationskoeffizienten" kann hier verzichtet werden, da sie in jedem Lehrbuch der Regressions- und Korrelationsanalyse nachgelesen werden können.[1)]

Die Aussage des einfachen Korrelationskoeffizienten läßt sich auf multiple Zusammenhänge analog übertragen. In diesem Fall gibt sein Quadrat, das multiple Bestimmtheitsmaß, den Anteil der Veränderung in der Zielgröße an, der durch die Veränderungen in allen in die Rechnung einbezogenen unabhängigen Variablen erklärt werden kann.

Während die multiplen Regressionskoeffizienten, wie bereits dargestellt,[2)] mit den partiellen Regressionskoeffizienten identisch sind, besteht zwischen der multiplen und der partiellen Korrelation - und damit auch zwischen den entsprechend bezeichneten Korrelationskoeffizienten - ein wesentlicher Unterschied. Der multiple Korrelationskoeffizient drückt die "Stärke" oder "Strammheit" des Zusammenhangs zwischen a l l e n in die Analyse einbezogenen Faktoren und der Zielgröße aus; der partielle Korrelationskoeffizient dagegen gibt den Grad der Verbundenheit zwischen e i n e r unabhängigen Variablen und der Zielgröße an, wenn in die Analyse mehrere unabhängige Veränderliche eingegangen sind, ihr Einfluß auf die Zielgröße aber ausgeschaltet wurde.[3)]

---

1) Vgl. z.B. E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 110 ff.; M. EZEKIEL u. K.A. FOX, a.a.O., S. 126 ff. u. S. 251 ff.; O. ANDERSON, a.a.O., S. 189 ff.; F.E. CROXTON u. D.J. COWDEN, a.a.O., S. 461 ff.

2) Vgl. S. 24 f.

3) Vgl. E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 20 f. u. S. 130 ff.; M. EZEKIEL u. K.A. FOX, a.a.O. S. 192 ff.; C.H. GOULDEN: Methods of Statistical Analysis. Second edition. New York und London (1952). S. 134 ff.

Durch die partielle Korrelation kann die innere Beziehungsstruktur zwischen mehreren Veränderlichen besser aufgedeckt werden, da mehrere gleichzeitig wirkende Ursachen, die eine Folgeerscheinung gemeinsam hervorbringen, isoliert in ihrem Einfluß auf die Zielgröße dargestellt werden können.[1] So kann eine einfache Korrelation zwischen zwei Variablen einen negativen Korrelationskoeffizienten zeigen, während die gleiche Korrelation unter Einbeziehung einer weiteren Größe einen positiven Wert für den multiplen Korrelationskoeffizienten ergeben kann, weil die später in die Analyse aufgenommene Variable mit der schon in der Untersuchung befindlichen unabhängigen Variablen eine so starke negative Korrelation aufweisen kann,[2] daß dadurch der vorher untersuchte Zusammenhang sich in einem negativen Korrelationskoeffizienten niederschlägt.[3]

In solchen Fällen ist es notwendig, partielle Korrelationen zwischen der Zielgröße und jeder unabhängigen Veränderlichen sowie zwischen den unabhängigen Variablen untereinander aufzustellen, bei denen jeweils der Einfluß der anderen Variablen ausgeschaltet wird, d.h., in denen die Größen der anderen konstant gehalten werden. Dadurch wird die reine stochastische Beziehung zwischen den jeweils zwei betrachteten Variablen wiedergegeben.[4]

---

1) E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 21.

2) Auf die Problematik von Interkorrelationen zwischen zwei oder mehreren unabhängigen Veränderlichen, üblicherweise als "Multikollinearität" bezeichnet, wird an anderer Stelle dieser Arbeit noch ausführlich eingegangen. Vgl. den Abschnitt 5.3.4 auf S. 111 ff.

3) Vgl. das Beispiel bei E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 21.

4) Ebenda.

3. Kapitel

AUFSTELLUNG EINER GESAMTBETRIEBLICHEN KOSTENFUNKTION

Um einen ersten Überblick über einzelne Probleme der Anwendung der Regressions- und Korrelationsanalyse in der Kostenrechnung zu geben, wird in diesem Kapitel auf Untersuchungen eingegangen, deren Ziel in der empirischen Ermittlung einer gesamtbetrieblichen Kostenkurve bestand. Dabei soll zunächst eine Aufzählung der in der Literatur am häufigsten vorgebrachten Kritikpunkte genügen, um zu zeigen, daß diese Probleme - wenn auch teilweise in abgewandelter Form und geringerer Intensität - bei den im Rahmen dieser Arbeit untersuchten Zusammenhängen wieder auftreten.

Von den in den dreißiger und vierziger Jahren besonders in den Vereinigten Staaten von Amerika zahlreich durchgeführten empirischen Kostenuntersuchungen sind wohl die Arbeiten von Joel DEAN am bekanntesten geworden, der unter anderem in einer Strumpfwirkerei und in einer Transmissionsriemenfabrik empirische Gesamtkostenkurven abzuleiten versuchte.[1)]

Darüber hinaus wurden zwei groß angelegte Untersuchungen von Th. YNTEMA und anschließend von K.H. WYLIE und M. EZEKIEL durchgeführt, deren Ziel darin bestand, aus den Quartals-

---

1) J. DEAN: Statistical Cost Functions of a Hosiery Mill. "Studies in Business Administration". 11 (1941) 4, S. 5 ff.; J. DEAN: The Relation of Cost to Output for a Leather Belt Shop. "National Bureau of Economic Research - Technical Paper". No. 2. New York 1941. S. 3 ff.; J. DEAN: Statistical Cost Curves. "Journal of the American Statistical Association". 32 (1937) 197, S. 83 ff.

und Jahreszahlen der "U.S. Steel Corporation" eine approximierte Kostenkurve für die gesamte amerikanische Stahlindustrie aufzustellen.[1]

Diese Arbeiten haben - auch bis in die letzten Jahre - eine große Anzahl von Veröffentlichungen nach sich gezogen, in denen Ausgangspunkte, Vorgehensweisen und Ergebnisse kritisch untersucht werden. Deutlich hat sich gezeigt, daß die empirische Ermittlung einer g e s a m t-betrieblichen Kostenkurve äußerst problematisch ist. So ergeben sich beispielsweise Schwierigkeiten bei der Festlegung der Untersuchungsperioden, der Abgrenzung des Datenmaterials, der Bestimmung der unabhängigen Variablen, der Kostenerfassung und Kostenbereinigung und bei der Auswertung und Interpretation der Untersuchungsergebnisse.[2]

Insbesondere die im Zusammenhang mit der Bereinigung der Kosten auftretenden Probleme lassen es nicht zulässig erscheinen, auf Grund der in den zitierten Untersuchungen gewonnenen Ergebnisse die These vom linearen Verlauf der betrieblichen Gesamtkostenkurve zu unterstützen, da das benutzte Datenmaterial vielfach so geartet war, daß eine Verzerrung in Richtung auf die Linearität der Kostenkurve zwangsläufig auftreten mußte.[3] So wurden wichtige Teile

---

1) K.H. WYLIE u.M. EZEKIEL: The Cost Curve for Steel Production. "The Journal of Political Economy". 48 (1940) 6, S. 777 ff.; M. EZEKIEL u. K.H. WYLIE: Cost Functions for the Steel Industry. "Journal of the American Statistical Association". 36 (1941) 213. S. 91 ff.

2) Einzelheiten vgl. bei K. WEBER: Die Aussagefähigkeit..., a.a.O., S. 47 ff.; R. RUGGLES: The Concept of Linear Total Cost-Output Regressions. "The American Economic Review". 31 (1941) 2, S. 332 ff.; A.A. WALTERS: Expectations and the Regression Fallacy in Estimating Cost Functions. "The Review of Economics and Statistics". 42 (1960) 2, S. 210 ff.

3) K. WEBER: Die Aussagefähigkeit..., a.a.O., S. 74.; E. STRAUS: Cost Accounting and Statistical Cost Functions. "The American Economic Review". 35 (1945) 3, S. 430 ff.

der Kosten von vornherein als konstant pro Ausbringungseinheit angesehen, was einer linearen Gesamtkostenkurve entspricht. Andererseits lassen die bei allen Untersuchungen ermittelten linearen Kostenkurven - trotz der erwähnten Vorbehalte - erkennen, daß diese Kostenfunktionen schwerlich denen entsprechen können, auf denen die ältere Theorie aufbaut[1] (ertragsgesetzlich verlaufende Kostenkurven).

Bei den schon genannten Untersuchungen von YNTEMA, WYLIE und EZEKIEL ist besonders kritisch anzumerken, daß die Kostenkurve praktisch für einen gesamten Industriezweig ermittelt wurde und die einzelnen Zusammenhänge durch diese Globalanalyse verschleiert wurden. Da die Betriebe verschiedene Produkte herstellten und mit unterschiedlichen Kapazitätsauslastungen arbeiteten, war die Bildung von Durchschnittswerten problematisch.[2]

Sämtliche angeführten Untersuchungen basieren auf Kapazitätsausnutzungsgraden, die beträchtlich weniger streuen als die technisch mögliche Kapazitätsschwankung. Sie können demnach - wenn überhaupt - Anspruch auf Gültigkeit lediglich für den stark eingeengten untersuchten Beschäftigungsbereich erheben und lassen keine Aussagen über höhere oder niedrigere Auslastungsgrade zu, weil eine Extrapolation der gewonnenen Ergebnisse auf diese Beschäftigungshöhen problematisch ist.[3] Damit ist die Konstruktion einer

---

1) E. GUTENBERG: Grundlagen der Betriebswirtschaftslehre. 1. Band: Die Produktion. 17. Aufl. Berlin/Heidelberg und New York 1970. S. 381.

2) Vgl. M. EZEKIEL u. K.H. WYLIE, a.a.O., S. 92.; C.A. SMITH: The Cost-Output Relation for the U.S. Steel Corporation. "The Review of Economic Statistics". 24 (1942) 4, S. 166 ff.

3) Auf die Problematik der Extrapolation der Regressionsfunktion wird in dieser Arbeit noch näher eingegangen. Vgl. z.B. S. 50 ff. u. S. 89 ff.

gesamten Kostenkurve von der Beschäftigung Null bis zur Maximalbeschäftigung auf Grund des in den zitierten Untersuchungen verwendeten Datenmaterials unmöglich.[1)]

Zur Frage der Linearität der Kostenkurve ist weiterhin festzustellen, daß auch geringfügige Abweichungen von der Linearität in den Gesamtkosten schon zu einer beträchtlichen Krümmung in der Grenzkostenkurve führen können, wie APEL nachgewiesen hat.[2)] Diese denkbaren geringfügigen Abweichungen werden jedoch bei Aufstellung der Gesamtkostenkurve infolge von Verzerrungen des Ausgangsmaterials, zum Beispiel durch die üblicherweise angewandte Form linearer Abschreibung, eliminiert.[3)]

Auch wird von APEL der Vorwurf erhoben, kleine, nicht repräsentative Stichproben gewählt zu haben, die den Ergebnissen keine statistische Sicherheit gewähren.[4)] Auf dieses Problem ist in dieser Arbeit noch ausführlich einzugehen.[5)]

Probleme der Betriebsgrößenänderung und technischer Verfahrensänderungen (changes in size, technological changes) werden von STAEHLE angesprochen.[6)] Sie finden in den genannten Arbeiten von DEAN, YNTEMA, WYLIE und EZEKIEL kaum Berücksichtigung. Andere Autoren versuchen dieses Problem zu umgehen, indem sie Betriebe und Untersuchungsperioden so

---

1) Vgl. H. APEL: Marginal Cost Constancy and its Implications. "The American Economic Review". 38 (1948) 5, S. 870 ff., hier S. 872.

2) Ebenda. S. 873.

3) Ebenda. Vgl. auch S. 33 dieser Arbeit.

4) Ebenda. S. 874.

5) Vgl. den Abschnitt 5.6, S. 141 ff.

6) H. STAEHLE: The Measurement of Statistical Cost Functions: An Appraisal of some Recent Contributions. "The American Economic Review". 32 (1942) 2, S. 321 ff.

wählen, daß keine Betriebsgrößen- und Verfahrensänderungen auftreten. Dadurch wird jedoch der praktische Wert derartiger Untersuchungen erheblich gemindert.[1)]

Schon aus den vereinzelten hier angeführten Kritikpunkten geht hervor, daß die regressionsanalytische Ermittlung einer gesamtbetrieblichen Kostenfunktion so zahlreiche und vielschichtige Probleme mit sich bringt, daß ihre Ergebnisse nur mit äußerster Vorsicht verwendet werden können. Allerdings wird von vielen Autoren anerkannt, daß diese Untersuchungen dazu beigetragen haben, das Bewußtsein für die Komplexität des Themas zu erhöhen,[2)] die Relevanz der Informationen für Kostenplanung, Kostenkontrolle und Preispolitik zu betonen[3)] und - was nach Meinung des Verfassers sehr wichtig ist - "ein gutes Bild über die zur Durchführung empirischer Kostenuntersuchungen verfügbaren und praktisch verwendeten Verfahren [zu] vermitteln".[4)]

Besonders deutlich hat sich gezeigt, daß neben den Problemen, die jede empirische Kostenuntersuchung mit sich bringt - gleichgültig, mit welchen Methoden sie durchgeführt wird und mit welchen Problemen der Kostenrechnung sie sich befaßt[5)] - viele Schwierigkeiten und Ungenauigkeiten deshalb entstanden, weil versucht wurde, durch eine Globalanalyse eine Kostenfunktion für ein so komplexes Gebilde wie einen gesamten Betrieb oder gar einen ganzen Industriezweig auf-

---

1) H. STAEHLE, a.a.O., S. 321 ff.

2) Ebenda. S. 332.

3) J. DEAN: Statistical Cost Curves, a.a.O., S. 89.

4) K. WEBER: Die Aussagefähigkeit..., a.a.O., S. 74.

5) So ergeben sich auch im Rahmen dieser Arbeit z.B. Probleme der Abgrenzung der Beobachtungsperioden und des -materials, der Größe des Stichprobenumfangs, der Wahl der Funktionsform, der Extrapolation der Regressionslinie usw.

zustellen. Viele dieser Probleme werden gegenstandslos oder doch zumindest leichter lösbar, wenn man die Regressions- und Korrelationsanalyse nicht dermaßen global auf d i e Kosten anwendet, sondern wenn man einzelne, eng begrenzte Teilbereiche innerhalb der betrieblichen Kostenrechnung oder einzelne Kostenarten betrachtet und hier durch Anwendung der Regressionsanalyse - insbesondere in ihrer multiplen Form - Aufteilungen, Verrechnungen, Planungen und Kontrollen von Kosten vornimmt. Das soll in den folgenden Kapiteln - vielfach anhand von Beispielen - gezeigt werden.

4. Kapitel

ERFASSUNG UND VERRECHNUNG VON KOSTEN

4.1 Einführung

Die Möglichkeiten zur Anwendung der Regressions- und Korrelationsanalyse auf einzelne Probleme der Kostenrechnung werden in den folgenden Kapiteln dargestellt. Gemäß der im Abschnitt 1.2 als Rahmen dieser Arbeit vorgegebenen Zielsetzung soll zunächst auf die Erfassung und Verrechnung von Kosten für die Perioden- und Stückerfolgsrechnung eingegangen werden, wobei Probleme der Kostenarten-, Kostenstellen- und Kostenträgerrechnung angesprochen werden. Im Abschnitt 4.2 soll anhand eines Beispiels einer retrograden Materialverbrauchserfassung das Prinzip der regressionsanalytischen Kostenschätzung näher erläutert werden.

In dem folgenden Abschnitt 4.3 wird der Versuch beschrieben, wie Gemeinkosten mit Hilfe der multiplen Regressionsanalyse auf Kostenstellen und Kostenträger verteilt werden können.

Der letzte Abschnitt 4.4 dieses Kapitels schließlich zeigt die Anwendung der Korrelationsanalyse zur Bestimmung einer geeigneten Schlüsselgröße für die Gemeinkostenverrechnung, wobei die Höhe des Korrelationskoeffizienten für die betrachteten Abhängigkeiten das Kriterium für die Wahl der besten Schlüsselgröße bildet.

4.2 Erfassung von Gemeinkostenmaterial im Rahmen der Kostenartenrechnung

In der Kostenartenrechnung wendet man üblicherweise je nach der betrieblichen Bedeutung und dem Wert eines Mate-

rials unterschiedliche Methoden der Verbrauchsermittlung an. Für wertvolle und wichtige Materialarten wählt man in der Regel eine Einzelerfassung der Lagerabgänge durch Materialentnahmescheine, während für geringwertige Materialien häufig eine summarische Abgangserfassung durch sog. "Rückrechnung" erfolgt,[1] da eine Einzelerfassung oder auch eine Abgangserfassung durch Bestandsvergleich einen im Verhältnis zum Wert des Materials zu hohen Aufwand erfordern würde. "Hierbei multipliziert man die Erzeugungs- oder Ablieferungsmenge einer Periode mit den aus Erfahrung, Verbrauchsanalysen oder Stücklisten stammenden Soll-Verbrauchsmengen je Erzeugniseinheit",[2] die dann als tatsächlicher Verbrauch in der Kostenstellen- und Kostenträgerrechnung Verwendung finden. Da man bei diesem Verfahren den Verbrauch auf Grund der Produktion oder anderer Ausgangsgrundlagen - wie Einzelmaterial oder Einzellöhne - zurückrechnet, wird es üblicherweise als "retrograde Materialverbrauchsermittlung" bezeichnet.[3]

Dieses Verfahren bietet gegenüber der Einzelerfassung den Vorteil der praktischen Brauchbarkeit und hohen Wirtschaftlichkeit.[4] Die Schwierigkeiten bestehen in einer möglichst exakten Ermittlung der Soll-Verbrauchsmengen je Erzeugniseinheit, die den Ausgangspunkt für die Verbrauchsermittlung bilden. Sie können auf Grund von statistisch gewonnenen Erfahrungssätzen, aus Stücklisten oder durch Verbrauchsanalysen ermittelt werden.[5]

---

1) Vgl. H. NORDEN u. F. WILLE: Der Betriebsabrechnungsbogen. Die Kostenstellenrechnung in der Praxis. 10., aktualis. u. erw. Aufl. Stuttgart 1965. S. 64 f.

2) Ebenda. S. 64.

3) E. KOSIOL: Kalkulatorische Buchhaltung, (Betriebsbuchhaltung). Systematische Darstellung der Betriebsabrechnung und der kurzfristigen Erfolgsrechnung. 5., durchges. Aufl. Wiesbaden 1953. S. 161.

4) Ebenda.

5) Ebenda.

Wie kann nun eine solche Gewinnung von statistischen Erfahrungssätzen erfolgen? An dieser Stelle zeichnet sich eine Anwendungsmöglichkeit für die Regressionsanalyse ab.

Wir setzen den Fall, daß ein bestimmtes geringwertiges Material gleichzeitig für verschiedene Erzeugnisse verwendet wird, und zwar in unterschiedlichen, aber bisher nicht exakt bekannten Mengen. Dann besteht die Möglichkeit, für einen bestimmten festgelegten Zeitraum sowohl den gesamten Materialverbrauch als auch die Anzahl der Produktionsmengen einzelner Produktarten, für die das betreffende Material verbraucht wurde, zu beobachten und in Listen festzuhalten. Nehmen wir an, daß das interessierende Material für drei Produkte in den Perioden $i = 1, 2, \ldots, n$ in den Mengen x_{1i}, x_{2i} und x_{3i} benötigt wurde, so erhalten wir folgende Tabelle:

Periode	y_i	x_{1i}	x_{2i}	x_{3i}
1	y_1	x_{11}	x_{21}	x_{31}
2	y_2	x_{12}	x_{22}	x_{32}
3	y_3	x_{13}	x_{23}	x_{33}
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
i	y_i	x_{1i}	x_{2i}	x_{3i}
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
n	y_n	x_{1n}	x_{2n}	x_{3n}

Tabelle 4.1: Tabelle der beobachteten Werte für die abhängige Variable y (Verbrauchsmengen) und die unabhängigen Variablen x_1, x_2 und x_3 (Produktionsmengen) für die Beobachtungsperioden $i = 1, 2, \ldots, n$.

In diesem Schema bezeichnet also y_i die beobachtete Menge der abhängigen Variablen, d.h. die beobachtete Verbrauchsmenge des Materials in der Beobachtungsperiode i und x_{1i} (x_{2i}; x_{3i}) die beobachteten Werte der unabhängigen Variablen, d.h. die produzierten Mengen der Güter 1 bzw. 2 bzw. 3 in der Beobachtungsperiode i.

Auf Grund dieser Daten ist es möglich, eine multiple Regressionsgleichung der Form

$$y' = a_{y.123} + b_{y1.23}x_1 + b_{y2.13}x_2 + b_{y3.12}x_3 \qquad (4.1)$$

zu bilden und nach einem im 2. Kapitel beschriebenen Verfahren die Regressionsparameter zu errechnen.[1] Diese Gleichung gibt den Materialverbrauch in Abhängigkeit von unterschiedlich hergestellten Mengen der Produkte 1 bis 3 wieder.

Die partiellen Regressionskoeffizienten $b_{y1.23}$, $b_{y2.13}$ und $b_{y3.12}$ sagen dabei aus, wie hoch der Materialverbrauch für eine Einheit des betreffenden Produktes im Durchschnitt war. Nehmen wir an, wir erhalten für $b_{y1.23}$ den Wert 0,2, so können wir sagen, daß für eine produzierte Einheit des Gutes 1 im Durchschnitt 0,2 Einheiten des untersuchten Materials verbraucht wurden und damit auch als Verbrauchsschätzwert in der Kostenrechnung anzusetzen sind. In Zukunft ist es dann nicht mehr erforderlich, Verbrauchsmessungen vorzunehmen, da sich der gesamte Materialverbrauch durch Einsetzen der hergestellten Mengen der Produkte 1, 2 und 3 in die Regressionsgleichung näherungsweise bestimmen läßt. Zur Verdeutlichung mag das folgende Beispiel dienen:

Wir nehmen an, wir haben aus den Beobachtungswerten y_i, x_{1i}, x_{2i} und x_{3i} eine Regressionsgleichung der Form

---

1) Zur praktischen Berechnung vgl. 9. Anhang I: Rechnerische Durchführung von Regressions- und Korrelationsanalysen, S. 252 ff.

$$y' = 0,1 + 0,2\ x_1 + 1,1\ x_2 + 0,9\ x_3 \qquad (4.2)$$

erhalten. Im beobachteten Zeitraum wurden 30 Einheiten des Produktes 1, 20 Einheiten des Produktes 2 und 50 Einheiten des Produktes 3 hergestellt. Durch Einsetzen dieser Werte in die Regressionsgleichung (4.2) erhält man:

$$y' = 73,1,$$

d.h. 73,1 Mengeneinheiten des Materials sind in der Kostenrechnung als Materialverbrauch anzusetzen.[1)]

Voraussetzung für die sinnvolle Anwendung dieses Verfahrens ist, daß die Produktionsverhältnisse für einen längeren Zeitraum einigermaßen stabil bleiben, d.h. daß die in die Analyse einbezogenen Produkte nach dem gleichen Verfahren wie bisher auch in der Zukunft hergestellt werden. Wenn zwischenzeitlich zum Beispiel andere Maschinen oder andere Arbeitskräfte eingesetzt werden und dadurch die Möglichkeit von Veränderungen im Materialverbrauch gegeben ist, muß die weitere Gültigkeit der bisherigen Schätzwerte überprüft werden. Ergeben sich größere Abweichungen von den früher beobachteten Verbrauchsmengen, so ist zweckmäßigerweise eine neue Regressionsanalyse mit den veränderten Beobachtungswerten durchzuführen.[2)]

Dagegen behalten die partiellen Regressionskoeffizienten ihre Gültigkeit, wenn ein Produkt, das bisher als unabhängige Variable in der Analyse enthalten war, in Zukunft nicht mehr hergestellt wird. Diese Aussage baut auf der Tatsache auf, daß ein partieller Regressionskoeffizient diejenige Veränderung in der Zielgröße angibt, die auf die Veränderung des Wertes e i n e r Variablen um eine Ein-

---

1) Auf die Bedeutung der Regressionskonstanten $a_{y.123}$, deren Wert in Höhe von 0,1 in dem Schätzwert für $y' = 73,1$ enthalten ist, wird an anderer Stelle dieser Arbeit noch ausführlich eingegangen. Vgl. S. 50 ff.

2) Auf diese Problematik wird im Verlauf dieser Arbeit noch mehrmals einzugehen sein. Vgl. S. 86 ff. u. S. 168 ff.

heit zurückzuführen ist, wobei der Einfluß a l l e r a n d e r e n Variablen als konstant angesehen wird.[1] Voraussetzung für die Gültigkeit dieser Aussage ist, daß zwischen dem nicht mehr hergestellten Produkt und den weiterhin als unabhängige Veränderliche in der Analyse verbleibenden Produkten keine I n t e r k o r r e l a t i o n e n bestehen, d.h. daß die von jedem Produkt hergestellten Mengen voneinander unabhängig sind. Liegen zwischen den unabhängigen Variablen Interkorrelationen vor - diese Beziehung wird häufig auch als M u l t i k o l l i n e a r i t ä t bezeichnet - dann können die Berechnungen zu falschen Schätzwerten für die Verbrauchsmengen führen.[2] Allerdings ist das Auftreten von Multikollinearität in dem hier gewählten Beispiel recht unwahrscheinlich, denn es ist nicht einzusehen, weshalb sich der Materialverbrauch - zum Beispiel der Verbrauch an Farbe zum Lackieren einer Maschine vom Typ A - ändern sollte, wenn die Maschine vom Typ B nicht mehr hergestellt wird, dadurch eine unabhängige Variable aus der Gleichung ausscheidet und aus der bisherigen vierdimensionalen eine dreidimensionale Regression wird.

In der Praxis lassen sich für diese statistische Ermittlung von Soll-Verbrauchsmengen zahlreiche Beispiele finden. Das gesamte Gemeinkostenmaterial kann auf diese Weise verbrauchsmäßig erfaßt werden, so zum Beispiel Hilfs- und Betriebsstoffe (Öle, Fette, Schmierstoffe, Reinigungsmittel u.ä.), Heiz- oder Verpackungsmaterial, Farben und Lacke, um nur einige Beispiele zu nennen.

---

1) Vgl. die Eigenschaften der partiellen Regressionskoeffizienten, S. 25.

2) Das Problem der Multikollinearität wird im Abschnitt 5.3.4 ausführlich behandelt.
Vgl. S. 111ff.

Bei Lösungs- oder Reinigungsmitteln zum Beispiel könnte in einer Regressionsanalyse der Einfluß der Lufttemperatur auf den Verbrauch berücksichtigt werden, da hohe Temperaturen den Verdunstungsverlust ansteigen lassen.[1] Damit ließe sich entweder eine einfache Regressionsgleichung zwischen Verdunstung als abhängiger Variabler und der Lufttemperatur als unabhängiger Veränderlicher ermitteln oder eine multiple Regression zwischen dem Verbrauch an Lösungsmitteln und seinen verschiedenen Einflußfaktoren, wobei die Lufttemperatur in diesem Fall eine unabhängige Variable von mehreren in den Gleichungsansatz eingehenden unabhängigen Veränderlichen darstellen würde.

Neben der relativ einfachen Verbrauchsermittlung selbst bietet die Regressionsanalyse den Vorteil, die für den Verbrauch pro Einheit eines bestimmten Produktes ermittelten Schätzwerte (partielle Regressionskoeffizienten) und den Schätzwert für den Gesamtverbrauch (y'_i) statistisch absichern zu können, indem Konfidenzintervalle (Vertrauensgrenzen) gebildet werden. Auf diese Weise läßt sich die Genauigkeit der Schätzungen mit einer bestimmten statistischen Wahrscheinlichkeit für ein festzulegendes Schwankungsintervall angeben. Beispielsweise ist eine Aussage der Art möglich, daß der für den durchschnittlichen Materialverbrauch des Produktes 1 gefundene Schätzwert mit einer Wahrscheinlichkeit von 95% zwischen 0,19 und 0,21 liegt.[2]

---

1) F. WILLE: Plan- und Standardkostenrechnung. Leitfaden. (Betriebswirtschaftliche Bibliothek. Hrsg. W. Hasenack. Reihe A/VII). 2. Aufl. Essen 1963. S. 52.

2) Diese Aussage soll hier zunächst genügen. Wie Konfidenzintervalle für die Schätzwerte und die Regressionsparameter gebildet werden und von welchen Voraussetzungen man dabei ausgeht, wird ausführlich im Abschnitt 5.6 dieser Arbeit dargestellt. Vgl. S. 141 ff.

4.3 Zuordnung und Verrechnung von Kosten im Rahmen der Kostenstellen- und Kostenträgerrechnung

4.3.1 Ausgangspunkt der Untersuchung

In ähnlicher Weise, wie die einfache oder multiple Regressionsanalyse dazu dienen kann, Verbrauchsschätzungen auf Grund beobachteter Daten vorzunehmen, kann sie auch für die Verrechnung von Kosten auf Kostenstellen und Kostenträger herangezogen werden. Während jedoch bei der Verbrauchsschätzung der Schwerpunkt der Analyse auf der Bestimmung von Schätzwerten für die abhängige Variable (Materialverbrauchsmengen) lag, ist das Ziel der Regressionsanalyse im Zusammenhang mit Kostenzuordnungen oder -verrechnungen im wesentlichen darin zu sehen, den Einfluß einzelner Faktoren auf die Höhe der Kosten isoliert, d.h. getrennt von anderen Einflüssen, festzustellen. In Begriffen der Regressionsrechnung ausgedrückt kann man sagen, daß sich nach Durchführung der Rechnung das Interesse im Abschnitt 4.2 vor allem den Schätzwerten y'_i zuwandte, während in diesem Abschnitt der Interpretation der partiellen Regressionskoeffizienten $b_{y1.23...m}$, $b_{y2.13...m}$, ..., $b_{ym.12...m-1}$ eine besondere Bedeutung zukommt.

Der Charakter der Kostenrechnung besteht in der "Erfassung und Zurechnung der betrieblichen Kosten auf die Leistungen".[1] Soweit es sich bei den Kosten um Einzelkosten handelt, die für eine - und nur eine - Produktionseinheit (Kostenträger) entstehen, können sie unmittelbar der Leistungseinheit zugerechnet werden. Dagegen sind Gemeinkosten dadurch gekennzeichnet, daß sie für mehrere Produktions-

1) K. MELLEROWICZ: Kosten und Kostenrechnung. Bd. II: Verfahren, 1. Teil: Allgemeine Fragen der Kostenrechnung und Betriebsabrechnung. 4. umgearb. Aufl. Berlin 1966. S. 1.

einheiten gemeinsam anfallen und damit nicht einzeln erfaßt und nicht direkt dem einzelnen Kostenträger zugerechnet werden können. Für sie ist vielmehr eine besondere Verrechnung oder Umlage auf die Leistungseinheiten erforderlich.

Da Industriebetriebe in der Regel dadurch gekennzeichnet sind, daß unterschiedliche Produkte in verschiedenen Fertigungsstufen hergestellt werden und damit die einzelnen Betriebsabteilungen ungleichmäßig beanspruchen (heterogene mehrstufige Produktion), ist zur genaueren Erfassung der Gemeinkosten eine Kostenstellenrechnung erforderlich.[1] In ihr werden die Gemeinkosten am Ort ihrer Entstehung, den Kostenstellen, erfaßt und können dadurch den Kostenträgern besser zugerechnet werden.

Bisher wurde der Begriff "Gemeinkosten" ausschließlich für Kostenträgergemeinkosten gebraucht. Daneben kennt man auch die sog. "Kostenstellengemeinkosten", d.h. Kosten, die nicht für eine Kostenstelle einzeln zu erfassen und damit auch nicht unmittelbar zuzurechnen sind, weil sie nicht ausschließlich für eine Stelle entstanden sind. Auch sie können, ähnlich der Verrechnung der Kostenträgergemeinkosten auf die Kostenträger, nur mit Hilfe von Verteilungsschlüsseln den Kostenstellen zugerechnet bzw. auf die Kostenstellen umgelegt werden.[2]

---

1) Vgl. z.B. E. SCHMALENBACH: Kostenrechnung und Preispolitik. 8. erw. u. verb. Aufl., bearbeitet von R. BAUER. Köln und Opladen 1963. S. 351 f.

2) Vgl. z.B. H.-M. SCHÖNFELD: Kostenrechnung. (Betriebswirtschaftliche Studienbücher. Hrsg. H. Seischab. Reihe IV: Kaufm. Rechnungswesen). 2., um das Kap. F, V erw. Aufl., Stuttgart 1964. S. 45.; K. MELLEROWICZ, a.a.O., II, 1. S. 389.

4.3.2 Verrechnung von Gemeinkosten mit Hilfe der multiplen Regressionsanalyse

Im folgenden soll geprüft werden, inwieweit sich die Regressionsanalyse zur Lösung der Probleme, die bei der Verrechnung von Gemeinkosten auf Kostenstellen und Kostenträger entstehen, heranziehen läßt.

Zur Veranschaulichung wollen wir dabei von folgendem Beispiel ausgehen:[1] In einer Kostenstelle werden auf verschiedenen Maschinen, die elektrisch angetrieben werden, unterschiedliche Produkte hergestellt, wobei alle Maschinen an denselben Stromzähler angeschlossen sind. Es stellt sich die Frage, wie die monatlichen Stromkosten, die sich aus der abgelesenen Zahl der verbrauchten Kilowattstunden multipliziert mit dem kwh-Tarif ergeben, auf die verschiedenen Kostenträger aufzuteilen sind. Läßt man den Tarif pro kwh als Preis des Stromes einmal außer acht, um Schwankungen in den Kosten auf Grund von unterschiedlichen Tarifen im Zeitablauf auszuschalten, arbeitet man also mit konstanten Preisen, so braucht man von den Kosten lediglich die Mengenkomponente, den Verbrauch, zu betrachten. Diesen Verbrauch liest man während des Beobachtungsmonats an jedem Arbeitstag[2] am Stromzähler ab und zeichnet gleichzeitig die Produktionszahlen der auf den Maschinen hergestellten unterschiedlichen Erzeugnisse dieser Kostenstelle auf. So erhält man die für eine multiple Regressionsanalyse erforderlichen Beobachtungsreihen mit dem Stromverbrauch als abhängiger Variabler y und den Mengen der Produkte als unabhängigen Variablen x_1, x_2 und x_3.

---

1) In Anlehnung an G. MATT: Die schrittweise Regressionsanalyse und ihre Anwendungsmöglichkeit im kaufmännischen Bereich. "Ablauf- und Planungsforschung - Operational Research". 4 (1963) 4, S. 254 ff., hier S. 255.

2) Im folgenden Rechenbeispiel, vgl. Tabelle 4.2 auf S. 47, wurden 20 Arbeitstage für den beobachteten Monat zugrundegelegt.

Es läßt sich somit ein ähnliches Schema wie das in Tabelle 4.1 beschriebene und eine Regression der in Gleichung (4.1) wiedergegebenen Form aufstellen.[1)]

Zur Verdeutlichung wollen wir von dem folgenden Rechenbeispiel ausgehen:

Periode (Arbeitstag)	y_i	x_{1i}	x_{2i}	x_{3i}
1	10,1	11	4	7
2	17,8	10	15	3
3	13,5	8	10	8
4	10,3	15	1	10
5	16,8	9	14	11
6	14,7	13	9	5
7	10,8	10	5	9
8	15,9	13	10	7
9	15,4	15	8	7
10	19,6	12	14	12
11	14,8	10	10	10
12	14,0	8	11	6
13	16,9	16	8	13
14	16,8	14	10	9
15	17,1	11	12	10
16	18,0	9	14	11
17	16,4	9	12	12
18	13,7	13	7	8
19	15,1	11	10	9
20	19,8	16	12	11
$\sum$	307,5	233	196	178

Tabelle 4.2: Verbrauchsmengen y_i und Produktionsmengen x_{1i}, x_{2i} und x_{3i} von drei Erzeugnissen für 20 Arbeitstage im Beobachtungsmonat.

Es wurde eine lineare multiple Regressionsgleichung der Form

$$y' = a_{y.123} + b_{y1.23}x_1 + b_{y2.13}x_2 + b_{y3.12}x_3$$

1) Vgl. S. 39 f.

zugrundegelegt. Die nach der Methode der kleinsten Quadrate ermittelte Regressionsfunktion ergibt folgende Werte:[1]

$$y' = - 0{,}023 + 0{,}5295\ x_1 + 0{,}7960\ x_2 + 0{,}1607\ x_3. \quad (4.3)$$

Diese Gleichung sagt aus, daß für jede hergestellte Einheit von x_1 0,5295, für jede Einheit von x_2 0,7960 und für jede Einheit von x_3 0,1607 kwh Strom im Durchschnitt verbraucht wurden. Mit diesen Werten sind demnach die Produkte 1, 2 und 3 im Rahmen der Kostenträgerrechnung pro gefertigte Einheit zu belasten.

Die Anwendungsmöglichkeit der multiplen Regressionsanalyse im Rahmen der Kostenträgerrechnung ist selbstverständlich nicht auf die Verteilung von Stromkosten oder Stromverbrauch beschränkt. Überall, wo Gemeinkosten entstehen, die auf Grund unterschiedlicher Produktionsleistung in ihrer Höhe schwanken, lassen sich die auf eine Produkteinheit entfallenden Kosten oder - bei konstanten Preisen - die Verbrauchsmengen durch die Bestimmung der partiellen Regressionskoeffizienten ermitteln. Man denke zum Beispiel an Energiekosten, wie die Kosten für Gas, Wasser, Kohle, Öl, Dampf usw. oder an bestimmte Arten von Gemeinkostenmaterial wie Heiz- oder Verpackungsmaterial.[2]

In ähnlicher Weise, wie in dem angeführten Beispiel die Verrechnung von Kostenträgergemeinkosten auf die Kostenträger erfolgte, lassen sich auch Stellengemeinkosten auf verschiedene Kostenstellen aufteilen. Voraussetzung für die Anwendung der Regressionsanalyse ist allerdings unter anderem, daß die Werte der Verrechnungseinheiten, die die unabhängigen Variablen bilden, tatsächlich Änderungen unter-

---

1) Zur Ermittlung der Regressionsparameter vgl. 9. Anhang I über die rechnerische Durchführung von Regressions- und Korrelationsanalysen, S. 252 ff.

2) Vgl. S. 42.

worfen, d.h. variabel sind. So ist die Regressionsanalyse natürlich überflüssig, wenn man Heizkosten nach der installierten Rippenzahl der Heizkörper auf Kostenstellen verteilt, da lediglich die Heizkosten selbst, je nach Außentemperatur und entsprechender Regulierung der Heizleistung, von Periode zu Periode schwanken können, die Zahl der Heizrippen aber konstant bleibt.

Dagegen ist eine regressionsanalytische Aufteilung von Materialgemeinkosten auf verschiedene Kostenstellen in folgender Form möglich.[1] Die im Materialbereich - Lager, innerbetrieblicher Transport usw. - anfallenden Gemeinkosten können durch eine multiple Regressionsanalyse aufgeteilt werden, in der zum Beispiel

1. das Gewicht der Materialien in Tonnen,
2. das Volumen in Kubikmetern und
3. der zu den einzelnen Kostenstellen zurückgelegte Transportweg in Metern

die unabhängigen Veränderlichen bilden. Die partiellen Regressionskoeffizienten treffen dann die Aussage über den Kostenbetrag, der einer Kostenstelle pro Tonne, pro Kubikmeter Material oder pro zurückgelegtem Meter an Transportweg zu belasten ist.

Nehmen wir an, daß wir bei Durchführung der Rechnung, die auf den beobachteten Daten aufbaute, die in Gleichung (4.3) angegebenen partiellen Regressionskoeffizienten erhalten haben und daß eine bestimmte Kostenstelle im Abrechnungszeitraum Material im Gewicht von 20 Tonnen mit einem Volumen von 40 Kubikmetern erhalten hat und daß der durchschnittliche Transportweg vom Lager zu dieser Kostenstelle 50 Meter beträgt. Dann sähe die Rechnung zur Ermittlung der Materialgemeinkosten für die Kostenstelle im betrach-

1) Vgl. auch A. SCHNETTLER, a.a.O., S. 136.

teten Zeitraum unter Berücksichtigung der Regressionskonstanten $a_{y.123}$ folgendermaßen aus:

$$y' = -0,023 + 0,5295 \cdot 20 + 0,7960 \cdot 40 + 0,1607 \cdot 50$$

$$y' = 50,44.$$

Die betrachtete Kostenstelle ist folglich mit einem Kostenbetrag in Höhe von 50,44 Einheiten zu belasten.[1)]

4.3.3 Interpretation der Regressionskonstanten

Schwieriger als die Erklärung der Regressionskoeffizienten oder des Schätzwertes y' gestaltet sich die Interpretation des absoluten Gliedes der Regressionsgleichung. Auf den ersten Blick mag es - zumindest bei positivem Vorzeichen von $a_{y.123}$ - zweckmäßig und auch logisch erscheinen, diesen Wert - beziehen wir uns auf die Beispiele im Abschnitt 4.3.2 - als fixe Stromkosten der Kostenstelle oder als fixe Transportkosten des Betriebes zu deuten, die unabhängig von der unterschiedlichen Produktion in immer gleicher Höhe anfallen, bei der Untersuchung der Stromkosten zum Beispiel für die Raum- oder Maschinenbeleuchtung. Eine solche Auslegung ist jedoch nicht ohne weiteres zulässig, denn eine Regressionsgleichung hat grundsätzlich nur in dem Bereich Gültigkeit, in dem für die unabhängigen Variablen auch Beobachtungswerte vorliegen. Eine Verlängerung der Regressionsfunktion in Richtung auf den Ursprung des Koordinatensystems ist deshalb äußerst problematisch.[2)] Wenn in die Ana-

---

1) Das Ergebnis wurde auf zwei Stellen hinter dem Komma gerundet.

2) So heißt es bei M. EZEKIEL u. K.A. FOX, a.a.O., S. 138 im Zusammenhang mit einem Beispiel aus dem Bereich der Landwirtschaft, dessen Aussagen aber allgemeingültig sind: "Die Abhängigkeiten, die die Regressionslinie wiedergibt, gelten nicht notwendigerweise jenseits dieser Grenzen, ... Eine Extrapolation der Regressionslinie bis zu jenem Punkt [gemeint ist d e r Punkt des Koordinatensystems, bei dem der Wassereinsatz gleich null ist] ist ohne weitere Annahmen sinnlos." Bei diesem Zitat handelt es sich um eine eigene Übersetzung des Verfassers.

lyse für die unabhängigen Veränderlichen nur stark von null abweichende Werte einbezogen werden konnten, ist eine Interpretation des absoluten Gliedes der Regressionsgleichung als Fixkosten sinnlos.

In unserem im Abschnitt 4.3.2 entworfenen Beispiel der Stromkostenverrechnung hat sich für die Regressionskonstante sogar ein negativer Wert von a = -0,023 ergeben. In diesem Fall wäre eine Interpretation von a als Fixkosten oder Fixverbrauch schon auf Grund des negativen Vorzeichens unsinnig. Eine solche Aussage würde nämlich bedeuten, daß die Maschinen bei der Beschäftigung null, d.h. im Stillstand, nicht nur keine Kosten verursachen, sondern sogar Strom liefern.

Wie läßt sich aber das Auftreten eines negativen absoluten Gliedes der Regressionsfunktion deuten? Die Erklärung liegt in der Tatsache, daß erstens ohne nähere Prüfung des Datenmaterials eine l i n e a r e Regressionsfunktion unterstellt wurde[1] und zweitens nur ein bestimmter Beschäftigungsbereich untersucht werden konnte, der recht weit vom Nullpunkt des Koordinatensystems entfernt liegt.[2] In der Praxis wird man in den seltensten Fällen für die unabhängige(n) Variable(n) Werte beobachten können, die nahe an dem Wert 0 liegen, da das eine weit unter der für einen bestimmten Betrieb typischen Beschäftigungslage liegende Kapazitätsauslastung voraussetzen würde, die im allgemeinen

---

1) Auf das Problem der Wahl einer geeigneten Funktionsform wird an späterer Stelle dieser Arbeit ausführlich eingegangen. Vgl. Abschnitt 6.3.3, S. 200 ff.

2) Für x_1 liegt der niedrigste beobachtete Wert bei 8. Die relativ niedrigen Werte für x_2 mit 1 und x_3 mit 3 können an dieser Feststellung nichts ändern, da sie weit aus dem Rahmen der anderen Beobachtungswerte fallen und als "Ausreißer" angesehen werden können.

nicht realisiert wird.[1] Vielmehr wird man in diesen Fällen versuchen, mit der Bearbeitung der Aufträge zu warten, bis durch Hinzukommen neuer Aufträge eine einigermaßen "normale" Beschäftigungslage garantiert ist.[2]

Zur Verdeutlichung der Gefahr, die in der bedenkenlosen Interpretation der Regressionskonstanten als Fixkosten oder Fixverbrauch liegt, sollen die beiden folgenden Abbildungen beitragen. Beide Schaubilder beziehen sich auf einfache Abhängigkeiten, weil in diesem Fall die Darstellung durch die Verwendung eines zweidimensionalen Koordinatensystems erleichtert wird. Die anhand dieser Abbildungen getroffenen Aussagen gelten jedoch in gleicher Weise für mehrdimensionale Analysen, d.h. für Regressionsgleichungen mit mehr als einer unabhängigen Variablen.

Die Abbildung 4.1 zeigt, daß man bei einer Extrapolation der für einen bestimmten Beschäftigungsbereich gefundenen Beziehung zwischen dem Verbrauch und seiner Einflußgröße zu falschen Ergebnissen gelangen kann, wenn man das absolute Glied der Regressionsgleichung als fixen Verbrauch interpretiert, die "wahre" Beziehung zwischen Verbrauch und Beschäftigung aber der als "vermutlicher technologischer Verlauf" bezeichneten Kurve folgt. Statt des tatsächlichen Fixverbrauchs a' ergibt sich durch die Verlängerung der Regressionsgeraden bis zum Schnittpunkt mit der Ordinate der sinnlose, rechnerisch ermittelte negative Fixverbrauch a.

---

1) Auf die Bedeutung von Stillstands-, Stillegungs-, Wiederanlaufkosten usw. soll hier nicht eingegangen werden.

2) Vgl. hierzu auch Ph. LYLE: Regression Analysis of Production Costs and Factory Operations. Third edition, rev. by L.H.C. TIPPETT. Edinburgh und London 1957. S. 3.

Ebenso problematisch ist es - wie Abbildung 4.2 zeigt - eine durch den Koordinatenanfangspunkt gehende gekrümmte Linie durch eine durch den Ursprung gehende G e r a d e approximieren zu wollen. Auch hier gilt, daß sich im beobachteten begrenzten Beschäftigungsbereich Kurven vielfach durch eine Gerade recht gut approximieren lassen, daß dagegen eine Verlängerung dieser Geraden bis zum Schnittpunkt mit der Ordinate nicht sinnvoll ist.

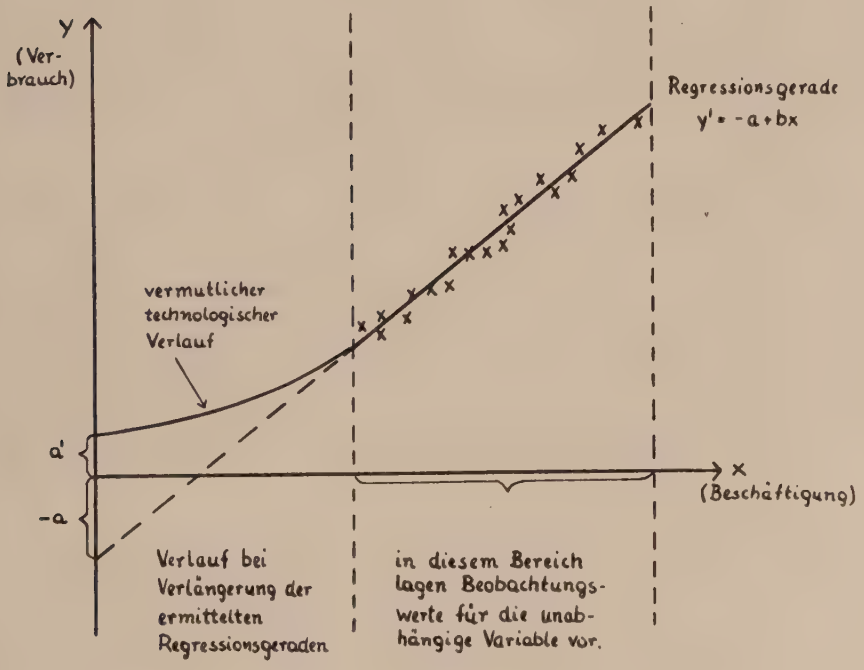


Abb. 4.1: Erklärung negativer Werte der Regressionskonstanten a bei linearem Verlauf der Regressionslinie im Beobachtungsbereich und nichtlinearem Verlauf in der Nähe der y-Achse.

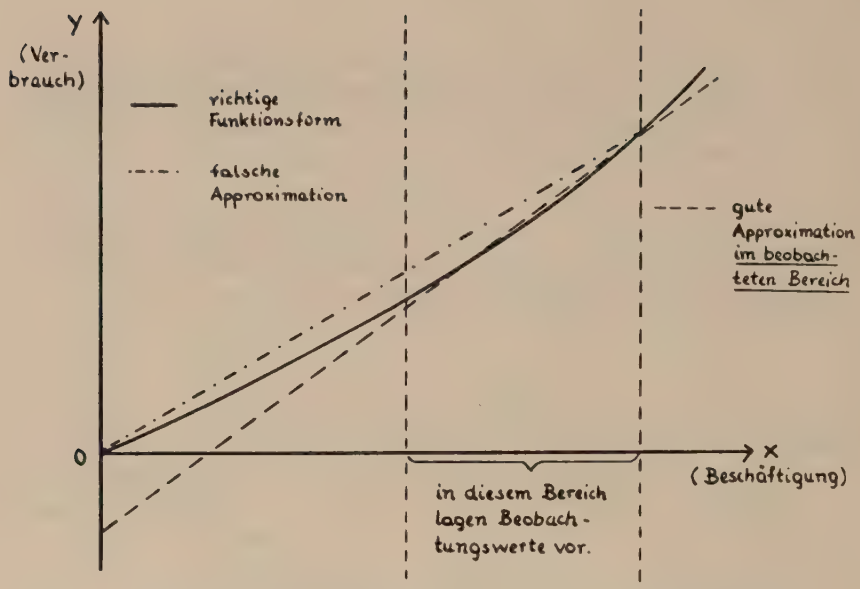


Abb. 4.2: Approximation einer Kurve durch eine Gerade.

Natürlich sind auch Situationen denkbar, in denen der für die Regressionskonstante rechnerisch ermittelte Wert ziemlich genau den fixen Kosten oder fixen Verbrauchsmengen entspricht. Jedoch lassen die möglichen Kurvenverläufe, wie sie in den Abbildungen 4.1 und 4.2 gezeigt wurden, erkennen, daß stets Vorsicht bei der Interpretation von Regressionskonstanten geboten ist.

Zusammenfassend kann festgestellt werden, daß der Regressionskonstanten a in aller Regel nur die Bedeutung eines auf Grund des Charakters der Ausgangsgleichung ermittelten Rechenwertes zukommt, der dem variablen Teil der Regressionsgleichung hinzuzufügen ist, um den genauen Schätzwert y' zu erhalten.

4.3.4 Vergleich der Regressionsergebnisse mit einer herkömmlichen Verrechnungsmethode

Um die Vorteilhaftigkeit der Regressionsschätzungen für das Rechenbeispiel der Stromkosten zu erläutern, sollen diese Werte denjenigen gegenübergestellt werden, die man nach einer herkömmlichen Verrechnungsmethode erhält. Dabei wird nach dem hier als "herkömmlich" bezeichneten Verfahren der auf eine Einheit eines Produktes entfallende Stromverbrauch nach der Gleichung[1)]

$$\text{Stromverbrauch pro hergestellte Einheit eines Produktes} = \frac{\text{Gesamtverbrauch}}{\text{Gesamtzahl aller hergestellten Produkte}}$$

ermittelt. Aus dieser Gleichung erhält man für den Stromverbrauch pro Einheit eines Produktes, wenn man die Werte aus Tabelle 4.2 auf Seite 47 übernimmt:

$$\frac{307,5}{607} = 0,5066.$$

Nach der herkömmlichen Methode ist demnach pro gefertigte Einheit eines Produktes - gleichgültig, ob es sich um Produkt 1, 2 oder 3 handelt - ein Stromverbrauch von 0,5066 kwh anzusetzen.

Um diesen Wert mit den Schätzwerten der Regressionsfunktion zu vergleichen, sollen am Beispiel der Beobachtungsperioden 1 bis 10 die Differenzen zwischen den Beobachtungswerten y_i und den Durchschnittswerten y_i' und y_i^+ gegenübergestellt werden. Mit y_i' werden die mit Hilfe der multiplen Regressionsgleichung ermittelten Schätzwerte bezeichnet, die y_i^+ repräsentieren die durch die herkömmliche Methode gewonnenen Schätzwerte für den Stromverbrauch.

---

1) Dieses Verfahren beruht auf der Annahme einer Gleichverteilung der unabhängigen Variablen.

Die Schätzwerte y_i' werden auf Grund der Regressionsgleichung

$$y' = 0{,}5295\ x_1 + 0{,}7960\ x_2 + 0{,}1607\ x_3,$$

die Durchschnittswerte y_i^+ nach der Gleichung

$$y^+ = 0{,}5066\ (x_1 + x_2 + x_3)$$

ermittelt.

Mit Hilfe des Differenzenvergleichs zwischen $|y_i - y_i'|$ und $|y_i - y_i^+|$ läßt sich erkennen, daß die durch die Regressionsfunktion ermittelten Schätzwerte den tatsächlichen Verbrauch weit besser wiedergeben als die einfachen Durchschnittssätze, da bis auf die Abweichung in Periode 5 alle Differenzen

Periode	y_i	y_i'	y_i^+	$\lvert y_i - y_i' \rvert$	$\lvert y_i - y_i^+ \rvert$
1	10,10	10,13	11,14	0,03	1,04
2	17,80	17,72	14,18	0,08	3,62
3	13,50	13,48	13,17	0,02	0,33
4	10,30	10,34	13,17	0,04	2,87
5	16,80	17,68	17,22	0,88	0,42
6	14,70	14,85	13,69	0,15	1,01
7	10,80	10,72	12,16	0,08	1,36
8	15,90	15,97	15,20	0,07	0,70
9	15,40	15,43	15,20	0,03	0,20
10	19,60	19,43	19,25	0,17	0,35

Tabelle 4.3: Differenzen zwischen beobachteten Werten y_i und berechneten Werten y_i' und y_i^+ für die abhängige Variable y.

zwischen Regresswert und Beobachtungswert wesentlich geringer sind als diejenigen zwischen "herkömmlichem Durchschnittswert" und Beobachtungswert.[1] Während man nämlich

1) Alle Werte wurden auf zwei Stellen nach dem Komma gerundet.

bei der herkömmlichen Methode lediglich mit e i n e m Durchschnittssatz für s ä m t l i c h e Produkte rechnen kann, erlaubt die Regressionsanalyse über die Ermittlung der partiellen Regressionskoeffizienten die Berücksichtigung der unterschiedlichen Maschinenbelastung der einzelnen Produkte. Die mit Hilfe der multiplen Regressionsrechnung ermittelten Schätzwerte lassen sich damit als "modifizierte Durchschnittswerte" auffassen, modifiziert nämlich im Hinblick auf die jeweils für eine Einheit des Produktes 1 oder 2 oder 3 entstehende Maschinenbelastung.[1)]

Darüber hinaus läßt der sehr hohe multiple Korrelationskoeffizient von $r_{y.123} = 0{,}9965$ erkennen, daß die beobachteten Werte durch die gewählte Form der Regression sehr gut repräsentiert werden. Mit Hilfe des Bestimmtheitsmaßes kann man aussagen, daß etwa 99,5% der Veränderungen im Stromverbrauch durch die Veränderungen in den Produktionszahlen der Güter 1, 2 und 3 erklärt werden. Testet man den multiplen Korrelationskoeffizienten anhand eines dafür entwickelten Nomogramms,[2)] so zeigt sich, daß bei einem Stichprobenumfang von $n = 20$, wie er in unserem Beispiel gewählt wurde, der tatsächliche (wahre) Korrelationskoeffizient der Grundgesamtheit[3)] mit einem Sicherheitsgrad von 95% mindestens den Wert 0,96 besitzt.[4)]

---

1) Über die Möglichkeit zur Bildung von Konfidenzintervallen für die Regreßwerte und die partiellen Regressionskoeffizienten vgl. Abschnitt 5.6, S. 141 ff.

2) Vgl. die Wiedergabe dieses Nomogramms im Anhang II: Einige wichtige Tabellen, S. 283ff., hier Nomogramm 8*, S. 291.

3) Auf das Problem, ob die im Rahmen eines Produktionsprozesses aufgezeichneten Beobachtungswerte als eine Stichprobe aus einer übergeordneten Grundgesamtheit oder als Grundgesamtheit selbst anzusehen sind, wird in dieser Arbeit noch näher einzugehen sein. Vgl. S. 142 ff.

4) Zur ausführlichen Darstellung des Tests von Korrelationskoeffizienten vgl. den folgenden Abschnitt 4.4.

Zusammenfassend kann festgestellt werden, daß die mit Hilfe einer multiplen Regressionsanalyse ermittelten Verbrauchsschätzwerte[1] gut für eine Verrechnung auf die Kostenträger geeignet sind. Die buchhalterische Erfassung der Abweichungen zwischen dem durch die Regressionsparameter geschätzten Verbrauch und dem tatsächlichen Verbrauch kann in der Kostenrechnung auf einem Konto "Materialverbrauchsabweichungen" erfolgen. Die dort gesammelten Differenzen können direkt auf ein Erfolgskonto in der Kontenklasse 9 (Abschlußkonten)[2] übernommen werden, ohne die Kostenstellen- und Kostenträgerrechnung zu durchlaufen.

4.4 Wahl der geeigneten Schlüsselgröße für die Gemeinkostenverrechnung

4.4.1 Problembeschreibung

Im vorhergehenden Abschnitt wurde gezeigt, auf welche Weise die multiple Regressionsanalyse bei der Verrechnung von Gemeinkosten auf Kostenstellen und Kostenträger angewendet werden kann. In den angeführten Beispielen wurden die Verrechnungseinheiten oder Bezugsgrößen als von vornherein gegeben angesehen. Bei der Verrechnung der Stromkosten auf unterschiedliche Produkte resultierte der verbrauchte Strom aus der Herstellung der genannten Produkte auf den Maschinen der betrachteten Kostenstelle. Es lag daher auf der

---

1) Die Berechnung der Verbrauchsschätzwerte y'_i und der Regressionsparameter bereitet schon unter Verwendung von programmierbaren Tischrechnern kaum Schwierigkeiten. Für umfangreiche und detailliertere Analysen stehen Programme für fast alle Typen von elektronischen Datenverarbeitungsanlagen zur Verfügung. Vgl. hierzu 9. Anhang I: Rechnerische Durchführung von Regressions- und Korrelationsanalysen. S. 252 ff.

2) Einteilung der Kontenklassen nach dem Gemeinschaftskontenrahmen der Industrie (GKR).

Hand, die Anzahl der gefertigten Produkte als Grundlage für die Aufteilung der Stromkosten bzw. des Stromverbrauchs zu wählen.

Jedoch existieren in der betrieblichen Praxis häufig Fälle, in denen eine solche Verrechnungseinheit für die Gemeinkosten nicht ohne weiteres gegeben ist, sondern in denen sich eine Vielzahl von möglichen Bezugsgrößen für die Verrechnung anbietet. In diesem Zusammenhang ist der Hinweis angebracht, daß nicht nur für die Verrechnung der Gemeinkosten auf die Kostenträger, sondern darüber hinaus generell für eine exakte Planung und Kontrolle der Gemeinkosten eine richtige Bezugsgrößenwahl unbedingte Voraussetzung ist,[1] so daß diesem Problem im Rahmen der Kostenrechnung eine große Bedeutung zukommt.

Um ein Beispiel für die Vielfältigkeit möglicher Bezugsgrößen zu geben, stelle man sich die Tätigkeiten in der Verwaltung oder im Vertrieb eines Unternehmens vor.[2] Für die Kosten aus dem Verkauf der Produkte lassen sich als Bezugsgröße[3] der Wert dieser verkauften Produkte oder die Anzahl der Kundenbesuche heranziehen, für die Kosten des Fertiglagers die Größe, das Gewicht oder die Anzahl der im Lager vorrätigen oder versandten Produkte. Für die Debitoren- oder die allgemeine Buchhaltung wären als Schlüsselgrößen für die Gemeinkostenverrechnung denkbar: die Anzahl

---

1) W. KILGER, a.a.O., S. 325 ff. In den folgenden Abschnitten wird ausführlich erläutert, was unter dem Begriff "richtige Bezugsgrößenwahl" zu verstehen ist.

2) Vgl. A. MATZ: Planung und Kontrolle von Kosten und Gewinn. Handbuch der Planungsrechnung. Deutsche Bearbeitung unter Mitwirkung von A. DEYHLE. (Bd. 7 der Schriftenreihe der AGPLAN, Hrsg. J. Fuchs, Ph. Kreuzer und K. Schwantag.) Wiesbaden 1964. S. 199 f.

3) Die Begriffe "Maßgröße", "Bezugsgröße", "Schlüsselgröße" und "Verrechnungseinheit" werden synonym verwendet. Anders D. SCHNEIDER: Kostentheorie und verursachungsgemäße Kostenrechnung. "Zeitschrift für handelswissenschaftliche Forschung", Neue Folge. 13 (1961) 11, S. 677 ff., hier S. 684.

der Kundenaufträge, der Rechnungspositionen, der Buchungsvorfälle oder ähnliche Größen.[1)]

Als Verrechnungseinheiten kommen also die verschiedensten Mengen- und Wertgrößen aus dem Gebiet des Produktionsaufwands (Arbeitslohn, Materialverbrauch, Zahl der Arbeits- oder Maschinenstunden), des Produktionsertrags (Anzahl der Produkte) und des Absatzes (Verkaufsmenge oder Verkaufserlös, Vertreterbesuche u.ä.) in Betracht.[2)]

Ausgehend von der üblicherweise als Hauptziel der Kostenrechnung angesehenen betrieblichen Periodenerfolgskontrolle, die die Verrechnung s ä m t l i c h e r Kosten auf Leistungseinheiten oder Verantwortungsbereiche bedingt,[3)] besteht das Problem darin, für die Verrechnung der Gemeinkosten eine Zuschlagsgrundlage zu finden, die dieser Zielsetzung am besten entspricht. In den folgenden Abschnitten soll daher untersucht werden, inwieweit die Regressions- und Korrelationsanalyse dazu beitragen kann, einen Lösungsansatz für das Problem der Gemeinkostenverrechnung zu liefern.

Bevor jedoch auf die empirische Bestimmung der Gemeinkostenfunktion eingegangen werden kann, erscheint es dem Verfasser notwendig, die kostenrechnerischen Grundlagen darzustellen, auf denen die Anwendung der Regressions- und Korrelationsanalyse in diesem Zusammenhang beruht.

4.4.2 Kostenverursachungsprinzip und Proportionalitätsprinzip

Das Kostenverursachungsprinzip, das man als das Fundamen-

---

1) Vgl. auch das Beispiel auf S. 49.
2) Vgl. K. KÄFER, a.a.O., S. 218.
3) Vgl. H. KOCH, a.a.O., S. 64.

talprinzip der Kostenrechnung bezeichnen kann und das auch die Grundlage für die folgenden Ausführungen bildet, besagt, daß die gesamten Kosten in der Weise auf die Leistungseinheiten verteilt werden sollen, daß jeder Einheit genau der Kostenanteil zugerechnet wird, den sie im Produktionsprozeß v e r u r s a c h t hat.[1] Dieses Prinzip ist auf die Verrechnung von Gemeinkosten nicht ohne weiteres anwendbar, da Gemeinkosten sich ex definitione nicht direkt, d.h. verursachungsgerecht auf Leistungseinheiten oder Verantwortungsbereiche zurechnen lassen, sondern mit Hilfe einer Bezugsgröße geschlüsselt werden müssen. Es muß deshalb ein Verteilungsprinzip gefunden werden, das über diese Schlüsselung eine operable Verrechnung der Gemeinkosten ermöglicht.

Von K. RUMMEL wurde aus dem Kostenverursachungsprinzip das sog. "Proportionalitätsprinzip" abgeleitet. Folgt man diesem Prinzip, dann sind die Gemeinkosten mit Hilfe derjenigen Bezugsgröße zu verteilen, bei der der relativ höchste Grad an Proportionalität mit dem zu verrechnenden Gemeinkostenbetrag gegeben ist.[2] Anders formuliert, man betrachtet die Zurechnung der Gemeinkosten über Schlüsselgrößen

---

1) Vgl. z.B.: K. RUMMEL: Einheitliche Kostenrechnung auf der Grundlage einer vorausgesetzten Proportionalität der Kosten zu betrieblichen Größen. 3. Aufl. Düsseldorf 1949. S. 17 u. S. 207.; M.R. LEHMANN: Industriekalkulation. (Betriebswirtschaftliche Bibliothek. Hrsg. W. Hasenack. Reihe A/VII). 5. Aufl. Essen 1964. S. 166 ff.; P. NOWAK: Kostenrechnungssysteme in der Industrie. (Bd. 21 der Veröffentlichungen der SCHMALENBACH-Gesellschaft). 2. erg. Aufl. Köln und Opladen 1961. S. 32.; E. HEINEN: Reformbedürftige Zuschlagskalkulation. "Zeitschrift für handelswissenschaftliche Forschung". 10 (1958) 1, S. 1 ff., hier S. 2.; K. MELLEROWICZ: Kosten und Kostenrechnung. Bd. II: Verfahren, 2. Teil: Kalkulation und Auswertung der Kostenrechnung und Betriebsabrechnung. 4. völlig umgearb. Aufl. Berlin 1968. S. 513.; K. KÄFER, a.a.O., S. 218.; H. KOCH, a.a.O., S. 66.

2) Ebenda. S. 68.; K. RUMMEL, a.a.O., S. 31 ff. u. S. 120 ff.; M.R. LEHMANN, a.a.O. S. 307.

nur dann als verursachungsgerecht, "wenn die Schlüsselgröße proportional, d.h. im gleichen Verhältnis wie die mengenmäßige Leistung variiert."[1)]

Diese Aussage gilt gleichermaßen für die Verteilung von Gemeinkosten auf Kostenstellen wie für die Zuteilung von Gemeinkosten auf Kostenträger.[2)] "Dabei bedeutet Proportionalität einmal, daß die Gemeinkostenfunktion linear ist, zum anderen, daß sie durch den Koordinatenanfangspunkt führt. Je nachdem nun, ob auch die fixen Gemeinkosten verrechnet werden oder nicht, wird mal der Gesichtspunkt der möglichst angenäherten Linearität, mal der Gesichtspunkt der größtmöglichen 'Kostenvariabilität' (d.h. des größtmöglichen Anteils der variablen Kosten) größeres Gewicht besitzen."[3)]

Es würde den Rahmen dieser Arbeit sprengen und ihrer Zielsetzung hinderlich sein, in aller Ausführlichkeit auf die in der Literatur wiedergegebene Kontroverse über die Eignung des Proportionalitätsprinzips für die Gemeinkostenschlüsselung einzugehen. Kritische Anmerkungen werden in einem späteren Abschnitt gebracht.[4)] An dieser Stelle soll zunächst die Eignung der Regressions- und Korrelationsanalyse zur Ermittlung von Gemeinkostenschlüsseln unter der Voraussetzung dargestellt werden, daß das Proportionalitätsprinzip als operables Verteilungsprinzip angesehen wird.

Dabei wird in diesem Zusammenhang noch einmal betont,[5)]

---

1) D. SCHNEIDER, Kostentheorie ..., a.a.O., S. 684.
2) M.R. LEHMANN, a.a.O., S. 167.
3) H. KOCH, a.a.O., S. 68.
4) Vgl. S. 74 ff.
5) Vgl. hierzu auch den Abschnitt 3.1, S. 31 ff.

daß der Verfasser die Regressions- und Korrelationsanalyse als mathematisches Hilfsmittel ansieht, das dem Kostenrechner wertvolle Anhaltspunkte für Entscheidungen liefern kann, das jedoch in aller Regel nicht in der Lage ist, Entscheidungen über die Richtigkeit betriebswirtschaftlicher (kostentheoretischer) Hypothesen zu treffen. Gegenstand der folgenden Untersuchung ist deshalb nicht eine Wertung des Proportionalitätsprinzips, sondern der Versuch, durch die Bestimmung von Korrelationskoeffizienten und ihre statistische Sicherung mit Hilfe von Elementen der Wahrscheinlichkeitsrechnung die Ermittlung von Gemeinkostenschlüsseln zu objektivieren. Mit Hilfe dieses mathematischen Modells kann eine Einschränkung des Ermessensspielraums gelingen, der in zur Zeit praktizierten Verfahren, z.B. bei der Bestimmung des Verteilungsschlüssels im Betriebsabrechnungsbogen, besteht. Da die dem Modell zugrunde liegenden Prämissen offengelegt werden müssen, können die zieladäquaten Modellmängel klar erkannt werden. Der Unternehmer wird sich der Problematik seiner Rechnung bewußt. Dies kann von den konventionellen Kostenrechnungsmethoden kaum behauptet werden.

4.4.3 Die empirische Bestimmung der Gemeinkostenfunktion

Geht man bei der Auswahl der für die Gemeinkostenverrechnung geeigneten Schlüsselgröße vom Kostenverursachungsprinzip und dem darauf aufbauenden Proportionalitätsprinzip aus, so bestehen grundsätzlich zwei Möglichkeiten der Bezugsgrößenwahl: die gedankliche Analyse (Bestimmung der Bezugsgröße nach rein theoretischen Gesichtspunkten) und statistische Verfahren, wobei hierunter hauptsächlich die Korrelationsanalyse zu verstehen ist.[1] Den statistischen Verfahren kommt in diesem Zusammenhang besondere Bedeutung zu, wenn die Einflußgrößen der Kostenverursachung rein gedanklich nicht mehr leicht zu überschauen sind.[2]

Bei der empirisch-statistischen Ermittlung der optimalen Bezugsgröße mit Hilfe der Korrelationsrechnung geht man in der Regel so vor, daß in einem Schaubild auf der Abszisse die über einen Zeitraum beobachteten Schlüsselmengen ihrer Größe nach aufgetragen werden, während auf der Ordinate die

---

1) Vgl. W. KILGER, a.a.O., S. 331. In den U.S.A. kennt man die gleiche Einteilung der Verfahren in "logical analysis" und "Correlation techniques". Vgl. "National Association of Cost Accountants (N.A.C.A.)-Bulletin." Section three. Research Series No. 6. Volume XXX, No. 20. June 1949. S. 1226., zitiert bei W. KILGER, a.a.O., S. 331.

2) Ebenda.

im Beobachtungszeitraum entstandenen Gemeinkostenbeträge angegeben werden.[1)]

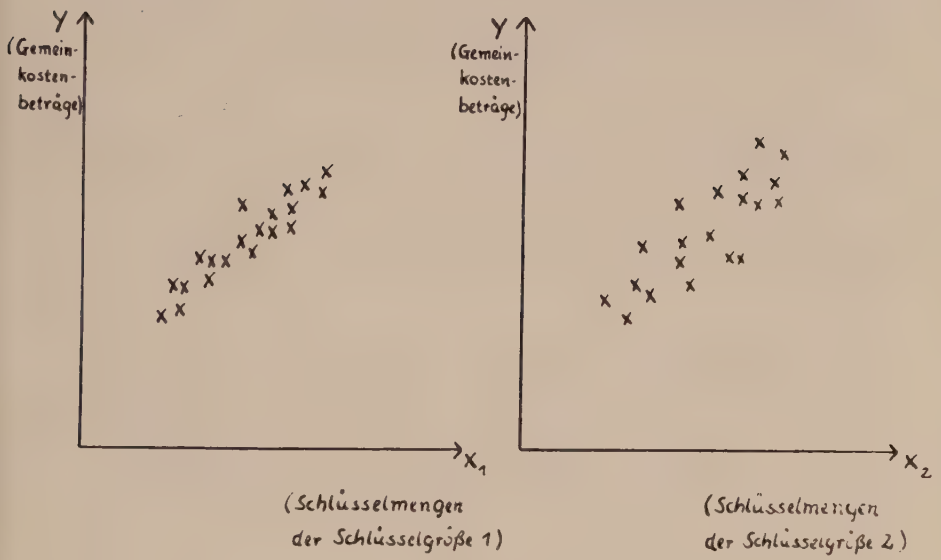


Abb. 4.3: Streupunktdiagramme für Beobachtungswerte aus Gemeinkostenbeträgen (y) und Schlüsselmengen der Schlüsselgrößen 1 (x_1) und 2 (x_2) für 20 Beobachtungsperioden.

---

1) Vgl. K. RUMMEL, a.a.O., S. 19 f. u. S. 180 ff.; K. KÄFER, a.a.O., S. 221 ff.; H. KOCH, a.a.O., S. 69.; E. HEINEN: Reformbedürftige Zuschlagskalkulation, a.a.O., S. 10 f.

Nachdem man auf diese Weise für die Beziehung zwischen Gemeinkosten und verschiedenen Schlüsselgrößen jeweils ein Streupunktdiagramm erhalten hat, läßt sich die Wahl der im Sinne des Proportionalitätsprinzips am besten geeigneten Schlüsselgröße entweder auf Grund einer graphischen Korrelationsanalyse oder auf Grund der rechnerischen Ermittlung des Korrelationskoeffizienten unter Verwendung der Methode der kleinsten Quadrate treffen.

Bei der graphischen Korrelationsanalyse wählt man als "Produktionsmaß jene Größe .., bei der sich die Kostenpunkte am deutlichsten einer Kurve, wenn möglich einer durch den Nullpunkt gehenden Geraden[1] anschmiegen".[2] Die Auswahl der optimalen Schlüsselgröße erfolgt bei diesem Verfahren also durch die Einzeichnung einer "Freihand-Kurve" in das Koordinatensystem, ein Verfahren, das zwar sehr einfach durchzuführen ist, jedoch den Nachteil einer mehr oder weniger großen Ungenauigkeit - je nach dem Charakter, d.h. der Streuung des vorliegenden Beobachtungsmaterials - mit sich bringt.[3] In den meisten Fällen ist daher die rechnerische Ermittlung der Korrelationskoeffizienten nach der Methode der kleinsten Quadrate zu empfehlen, denn nur auf diese Weise lassen sich q u a n t i t a t i v e Aussagen über die Höhe des Einflusses machen, den der einzelne betrachtete Einflußfaktor (Bezugsgröße) auf die Gemeinkosten ausübt.

Hierbei geht man so vor, daß man für verschiedene zur Auswahl stehende Schlüsselgrößen verschiedene Regressionen

---

1) Im Original kursiv gedruckt.

2) K. KÄFER, a.a.O., S. 220.

3) Vgl. auch W. KILGER, a.a.O., S. 331.

bildet, mit jeweils einer Schlüsselgröße als unabhängiger und den zu verteilenden Gemeinkosten als abhängiger Variabler. Für diese Regressionen ermittelt man jeweils den Korrelationskoeffizienten r_{yx} und wählt als optimale Bezugsgröße jene aus, die den höchsten Korrelationskoeffizienten aufweist, da die geforderte Proportionalität zwischen Gemeinkosten und Schlüsselgröße um so größer ist, je höher der Wert des Korrelationskoeffizienten ist.[1)]

Da die Berechnung von einfachen Korrelationskoeffizienten für lineare Abhängigkeiten keine besonderen Probleme aufweist,[2)] und da in der Praxis zahlreiche derartige Untersuchungen durchgeführt worden sind,[3)] erübrigt sich in diesem Rahmen die rechnerische Ermittlung von Korrelationskoeffizienten anhand konkreten Zahlenmaterials.

Auf der gleichen Ebene wie die Analyse von ROCKSTUHL liegen die Untersuchungen von GAYNOR, der ebenfalls anhand des einfachen Korrelationskoeffizienten für lineare Abhängigkeiten die "beste" Einflußgröße für Gemeinkosten

---

1) Vgl. die Ausführungen zum Korrelationskoeffizienten auf S. 27 ff.

2) Vgl. auch S. 28. Die dort angegebene Formel (2.19) läßt sich bei einfachen linearen Abhängigkeiten für Zwecke der praktischen Berechnung aus vorliegendem Beobachtungsmaterial umformen, so daß Rechengrößen verwendet werden können, die schon bei der Ermittlung der Regressionsparameter berechnet werden mußten.
Im übrigen vgl. hierzu die auf S. 29, Fußnote [1] angegebene Literatur und den Anhang I zur rechnerischen Durchführung von Regressions- und Korrelationsanalysen, S. 252 ff.

3) Vgl. z.B. die umfangreiche Untersuchung von Reparatur- und Instandhaltungskosten in einer Gesenkschmiede am am Beispiel von Schmiedehämmern, Schmiedepressen und Schmiedeöfen, die ausführlich beschrieben wird bei: J.P. ROCKSTUHL: Untersuchung über Möglichkeiten einer verursachungsgerechten Zuordnung der im betrieblichen Fertigungsablauf entstehenden Kosten, insbesondere der Restgemeinkosten. Diss. Aachen 1961. S. 42 ff.

sucht, allerdings nicht zur Verrechnung auf Kostenstellen oder Kostenträger, sondern zur Vorausschätzung der Kosten für Planungszwecke.[1] Hier wird lediglich ein anderes Ziel verfolgt, der Ausgangspunkt der Analyse und die Durchführung sind die gleichen wie bei der Ermittlung der optimalen Schlüsselgröße für die V e r r e c h n u n g der Gemeinkosten.

Auf ein besonderes Problem weist KÄFER hin, der bei der Untersuchung der Gemeinkosten der Kostenstelle Färberei eines Betriebes, der Verpackungsfolien herstellt, zu dem Ergebnis kam, daß sich die Stellengemeinkosten nicht nach dem Gewicht der Produkte in Kilogramm, nicht nach der Länge der Folien in laufenden Metern und auch nicht nach der Fläche in Quadratmetern oder nach der Anzahl der Bearbeitungsstunden in befriedigender Weise verrechnen ließen.[2] Aus diesem Grund unterteilte KÄFER die gesamten Gemeinkosten in Kostenartengruppen, die zu verschiedenen Maßgrößen proportional verlaufen könnten[3] und untersuchte dann den Zusammenhang zwischen diesen Gruppen von Kostenarten und den genannten vier Schlüsselgrößen.[4] Dadurch kommt er in seinem Beispiel zu dem Ergebnis, daß bei einer Zerlegung in stundenabhängige und flächenabhängige Kostenarten sich in Funktion der Fertigungsstunden und der Quadratmeterzahl eine genügend große Annäherung (d.h. ein genügend hoher Korrelationskoeffizient) ergibt, "so daß die Stelle zukünftig mit einem Stunden- und einem m^2-Satz abgerechnet wird".[5]

---

1) Vgl. E.W. GAYNOR, a.a.O., S. 1114 ff.

2) K. KÄFER, a.a.O., S. 221.

3) Vgl. hierzu auch: H. KOCH, a.a.O., S. 69.; J. FISCHER, O. HESS u. G. SEEBAUER: Buchführung und Kostenrechnung. 5. Aufl. Leipzig 1942. S. 327 ff.

4) Vgl. auch das Beispiel bei J.P. ROCKSTUHL, a.a.O., S.53.

5) K. KÄFER, a.a.O., S. 221.

Auch MATZ weist auf das Problem hin, daß im allgemeinen e i n e Leistungseinheit als Verschlüsselungsbasis genügt, jedoch auch Situationen auftreten, die mehrere Bezugsgrößen erfordern.[1] "So gibt es z.B. in der Kostenabteilung 'Verkauf' Kostenarten, die von der Anzahl der Aufträge, und andere, die vom Marktwert abhängen."[2] Auch hier ist es zur Ermittlung der optimalen Schlüsselgröße erforderlich, die gesamten Gemeinkosten der Kostenstelle in Gruppen aufzuteilen und für jede Gruppe getrennt den besten Schlüssel nach dem Kriterium der kleinsten Quadrate über die Bestimmung der Korrelationskoeffizienten herauszufinden.[3]

4.4.4 Test des Korrelationskoeffizienten

Um den auf der Grundlage eines bestimmten Datenmaterials errechneten Korrelationskoeffizienten zu prüfen, beziehen wir uns auf das in Tabelle 4.2 auf Seite 47 wiedergegebene Zahlenmaterial und bezeichnen nun die abhängige Variable y als Gemeinkostenbetrag einer Kostenstelle und die unabhängige Variable x_2 als Schlüsselgröße, so daß wir eine einfache Regression zwischen y und x_2 erhalten.[4] Die 20 Beobachtungswertepaare für y und x_2, die wir als Stichprobe aus einer

---

1) A. MATZ, a.a.O., S. 200.

2) Ebenda.

3) Wie die Zuordnung von einzelnen Kostenarten einer Kostenstelle zu bestimmten Kostenartengruppen erfolgen kann, hat ROCKSTUHL gezeigt. So bezieht er z.B. die Reparaturkosten auf Maschinenkosten und Lohnkosten und entscheidet auf Grund der Höhe der dafür errechneten Korrelationskoeffizienten, ob die Reparaturkosten der Kostenartengruppe "maschinenbedingte" oder der Kostenartengruppe "lohnbedingte" Gemeinkosten zuzuordnen sind. Vgl. J.P. ROCKSTUHL, a.a.O., S. 42 ff.

4) Der Einfachheit halber wird hier dasselbe Zahlenmaterial wie in Abschnitt 4.3.2 benutzt, wenn auch der Rechnungszweck ein ganz anderer ist.

übergeordneten Grundgesamtheit[1] interpretieren können, da wir lediglich 20 Wertepaare aus einer sehr viel größeren möglichen Gesamtheit ausgewählt haben,[2] liefern bei vorausgesetzter linearer Abhängigkeit einen Korrelationskoeffizienten von $r_{y2} = 0{,}8548$.

Es stellt sich auf Grund der Stichprobenerhebung die Frage, ob dieser Korrelationskoeffizient lediglich durch den Zufall erklärt werden kann, d.h. ob er nur zufällig von null abweicht, oder ob er "statistisch gesichert" ist, d.h. ob in der Grundgesamtheit tatsächlich eine Korrelation besteht. Um das zu prüfen, stellen wir eine Hypothese auf, die besagt, daß der wahre Korrelationskoeffizient der Grundgesamtheit - ϱ_{y2} - den Wert null besitzt. Diese Hypothese wird üblicherweise als "Nullhypothese" mit dem Symbol H_o bezeichnet.[3] Zur Überprüfung der Richtigkeit dieser Hypothese kann man eine Testgröße t ermitteln, die studentverteilt ist und für die bei einfacher Abhängigkeit[4] folgende Formel gilt:[5]

$$t = \frac{r}{\sqrt{1 - r^2}} \cdot \sqrt{n - 2} \quad . \qquad (4.4)$$

---

1) Unter dem Begriff "Grundgesamtheit" verstehen wir eine statistische Masse, d.h. eine Gesamtheit von Elementen, "die unter einem vom Untersuchungsziel her gesehenen Gesichtspunkt gleichartig sind". Vgl. H. KELLERER: Statistik im modernen Wirtschafts- und Sozialleben. (Bd. 103/104 von Rowohlts Deutscher Enzyklopädie, Sachgebiet Wirtschaftswissenschaften, Hrsg. E. Grassi). O.O. (1963). S. 11.

2) Vgl. dazu auch H.-M. SCHÖNFELD: Die Anwendung statistischer Methoden in der Kostenrechnung. (I). "Kostenrechnungspraxis. Zeitschrift für Betriebsabrechnung, Kostenrechnung und -planung". O.J. (1968) 6, S. 245 ff., hier S. 248.

3) Vgl. auch Kap. 6 (Kostenkontrolle).

4) D.h. bei einer Anzahl der Freiheitsgrade von $\nu = n - 2$.

5) Vgl. z.B. E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 225.; F.E. CROXTON u. D.J. COWDEN, a.a.O., S. 722. Die Studentverteilung gilt unter der Voraussetzung der Gültigkeit der Nullhypothese.

Da einfache Korrelationskoeffizienten nur Werte zwischen -1 und +1 annehmen können, folgen die Stichprobenkorrelationskoeffizienten nicht - auch nicht annähernd - einer Normalverteilung,[1] sondern genügen bei Gültigkeit der Nullhypothese einer Studentverteilung.[2]

Um die Nullhypothese $H_0: \varrho_{y2} = 0$ gegen die einseitige Alternativhypothese[3] $H_1: \varrho_{y2} > 0$ zu testen, errechnet man nach der Formel (4.4) auf Grund des gegebenen Datenmaterials den Testwert t und vergleicht ihn mit einem "kritischen Wert" t_c, der sich je nach gefordertem Sicherheitsgrad der Aussage aus der Tabelle der Studentverteilung ablesen läßt.[4] Wenn wir unterstellen, daß wir die Nullhypothese dann ablehnen wollen, wenn eine Wahrscheinlichkeit von mindestens 95% gegen ihre Richtigkeit spricht, d.h. wenn wir ein Fehlerrisiko von 5% bei unserer Entscheidung eingehen wollen, erhalten wir aus der Tabelle der Studentverteilung für $\nu = n - 2 = 20 - 2 = 18$ Freiheitsgrade einen kritischen Wert von $t_c = 1{,}734$. Das bedeutet, wenn wir für t einen Wert errechnen, der größer als t_c ist, können wir mit einem Sicherheitsgrad von mindestens 95% davon ausgehen, daß der wahre Korrelationskoeffizient ϱ_{y2} der Grundgesamtheit größer als null ist und daß wir den Stichprobenkorrelationskoeffizienten r_{y2} nicht lediglich zufällig erhalten haben.

---

1) Zur Anwendung der Normalverteilung vgl. auch Kap. 5 u. 6.

2) Vgl. E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 225 und S. 233. Kann man nicht von der Voraussetzung $H_0: \varrho_{y2} = 0$ ausgehen, muß die sog. "FISHERsche z-Transformation" angewendet werden. Vgl. dazu S. 73.

3) Man bezeichnet diesen Test als "einseitig", weil nur getestet wird, ob ϱ_{y2} g r ö ß e r als null ist, die Untersuchung sich also auf e i n e Seite der Verteilungsfunktion (vom Mittelwert aus gesehen) beschränkt.

4) Vgl. im Anhang II die Tabelle der Studentverteilung bei einseitiger Fragestellung. Tabelle 3*, S. 286.

Setzt man den für r_{y2} ermittelten Wert in die Formel (4.4) ein, so erhält man:

$$t = \frac{0,8548}{\sqrt{1 - 0,8548^2}} \cdot \sqrt{20 - 2} = 7,0.$$

Da das für unsere Stichprobe berechnete t = 7,0 größer ist als der kritische Wert $t_c = 1,734$, muß die Nullhypothese - der Korrelationskoeffizient der Grundgesamtheit sei null - abgelehnt werden.[1)]

Man kann demnach auf Grund des Stichprobenergebnisses davon ausgehen, daß der Korrelationskoeffizient ϱ_{y2} signifikant von null verschieden ist, was gleichbedeutend ist mit der Aussage, daß zwischen den Gemeinkosten und der betrachteten Schlüsselgröße 2 in der Grundgesamtheit eine Korrelation existiert. Ob diese Schlüsselgröße die beste aller möglichen ist, muß durch einen Vergleich zwischen den Korrelationskoeffizienten aller in Frage kommenden Schlüsselgrößen festgestellt werden. Beziehen wir uns wieder auf das Datenmaterial der Tabelle 4.2 auf Seite 47 und ermitteln auch für die Abhängigkeiten y von x_1 und y von x_3 die zugehörigen einfachen Korrelationskoeffizienten, so erhalten wir folgende Werte für Größe 1 und 3:

$$r_{y1} = 0,1227 \quad \text{und}$$

$$r_{y3} = 0,3004.$$

Für r_{y1} und r_{y3} ergeben sich so niedrige t-Werte, daß nicht auf Signifikanz dieses Stichprobenergebnisses geschlossen werden kann, d.h. die Nullhypothese kann in diesen Fällen

---

1) Auch bei einem Konfidenzniveau von 99 % mit einem kritischen Wert von $t_c = 2,552$ käme man zu demselben Schluß.

nicht abgelehnt werden. Es muß damit gerechnet werden, daß die Werte beider Korrelationskoeffizienten sich lediglich zufällig ergeben haben. Damit ist von den drei betrachteten Schlüsselgrößen die Größe 2 mit einem $r_{y2} = 0,8548$ für die Verrechnung der Gemeinkosten weitaus am besten geeignet.

Auf die Bestimmung von Konfidenzintervallen für die ermittelten Korrelationskoeffizienten kann verzichtet werden, da lediglich getestet werden sollte, inwieweit auf Grund des Stichprobenergebnisses auf eine Korrelation in der Grundgesamtheit geschlossen werden kann.[1)]

Neben der Berechnung der Testgröße t können Korrelationskoeffizienten auch anhand von Nomogrammen getestet werden, die auf der von R.A. FISHER entwickelten sog. "z-Transformation" zur Bestimmung von Vertrauensgrenzen für Korrelationskoeffizienten aufgebaut sind.[2)] Diese Tabellen basieren auf der Überlegung, daß, auch wenn man auf Grund des Stichprobenergebnisses nicht genaue Angaben über die wahre Korrelation in der Grundgesamtheit machen kann, man dennoch

---

1) Auch auf die Durchführung eines "F-Testes" soll an dieser Stelle verzichtet werden. Er führt zu entsprechenden Ergebnissen für das Bestimmtheitsmaß, vgl. z.B. E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 236 f. Zur ausführlichen Darstellung eines F-Testes vgl. den Abschnitt 6.3.3.3.2, S. 204 ff. in Verbindung mit der Tabelle 5* der F-Verteilung im Anhang II, S. 288.

2) Die von R.A. FISHER eingeführte Transformation der r-Werte in z-Werte ("z-Transformation") ist dann erforderlich, wenn nicht von der Nullhypothese $\varrho_{yx} = 0$ ausgegangen werden kann, da in diesem Fall die Verteilung der Stichprobenkorrelationskoeffizienten asymmetrisch ist. Zur FISHERschen z-Transformation vgl. z.B.: E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 234.; M. EZEKIEL u. K.A. FOX, a.a.O., S. 293.; R.A. FISHER: The General Sampling Distribution of the Multiple Correlation Coefficient. "Proceedings of the Royal Society", A, Vol. 121, (1928) S. 654 ff., zitiert bei M. EZEKIEL u. K.A. FOX, a.a.O., S. 295. Vgl. auch die Tabelle der z-Transformation im Anhang II, Tabelle 6*, S. 289.

einen Minimalwert für den Korrelationskoeffizienten der Grundgesamtheit schätzen kann, und zwar mit einem bestimmten quantifizierbaren Fehlerrisiko.[1]

Aus dem für einen Stichprobenumfang von n = 20 und einem Konfidenzniveau von 0,95 oder 95% aufgestellten Nomogramm[2] kann man entnehmen, daß für unser Beispiel[3] mit einem r_{y2} = 0,8548 der wahre Korrelationskoeffizient der Grundgesamtheit mindestens den Wert von 0,70 besitzt. Auch auf Grund dieser Tatsache läßt sich die Aussage bekräftigen, daß die Schlüsselgröße 2 gut als Verrechnungseinheit für die Gemeinkosten dieser untersuchten Kostenstelle geeignet ist.

4.4.5 Kritik der statistischen Ermittlung der optimalen Schlüsselgröße

4.4.5.1 Allgemeine Beurteilung

Bei der praktischen Ermittlung der optimalen Schlüsselgröße können sich dann Schwierigkeiten ergeben, wenn die für die Korrelationsanalyse benötigten Zahlenreihen in der für die Durchführung der Rechnung erforderlichen Genauigkeit nicht zur Verfügung stehen. Zwar liegen in der Regel Aufzeichnungen über die Gemeinkosten vor, dagegen werden häufig die Werte der zur Auswahl stehenden Bezugsgrößen nicht exakt für einen längeren Beobachtungszeitraum aufgeschrieben.[4] Für eine aussagefähige Korrelationsanalyse ist

---

1) M. EZEKIEL u. K.A. FOX, a.a.O., S. 293. (Sinngemäße Übersetzung des Verfassers).

2) Vgl. die Wiedergabe dieses Nomogramms im Anhang II, Nomogramm 7*, S. 290.

3) Vgl. S. 71 f.

4) Vgl. W. KILGER, a.a.O., S. 331 f.

jedoch unbedingte Voraussetzung, daß nicht nur für die Werte der abhängigen, sondern ebenso für die Werte der unabhängigen Variablen genügend genaue und genügend viele Aufzeichnungen vorhanden sind.

Weiter ist für diese Art der statistischen Analyse vorauszusetzen, daß die Werte der unabhängigen Variablen während der Beobachtungsperioden sich über einen möglichst großen Schwankungsbereich erstrecken. Je weniger die Beobachtungswerte für die unabhängige Variable streuen, um so größer ist die Gefahr von sog. "Streupunktballungen", d.h. von Ansammlungen der Beobachtungspunkte auf einer sehr kleinen Fläche im Koordinatensystem - vgl. Abbildung 4.4 - die die Ermittlung eines sinnvollen Korrelationskoeffizienten erschweren, wenn nicht unmöglich machen.

Abb. 4.4: Streupunktballung bei 20 Beobachtungswerten für die abhängige Variable y (Kosten) und die unabhängige Variable x (Beschäftigung).

Hat man auf Grund des vorliegenden Datenmaterials einen statistisch gesicherten Korrelationskoeffizienten ermitteln können und sich auf Grund seiner Höhe für eine bestimmte Schlüsselgröße als Verrechnungseinheit entschlossen, so bleibt die Frage, ob bei denkbaren zukünftigen Beschäftigungsänderungen, die größer sind als die in die Korrelationsanalyse eingegangenen - d.h. bei Ausdehnung der beobachteten Abszissenwerte auf höhere oder niedrigere Werte,

als sie bisher beobachtet werden konnten - die bisherige Schlüsselgröße auch die beste Größe bleibt oder ob nicht durch das neue Datenmaterial sich eine andere Bezugsgröße durch einen höheren Korrelationskoeffizienten als besser für die Gemeinkostenverrechnung erweist. Hier treten die schon erwähnten Probleme der Extrapolation, d.h. der Unterstellung der Gültigkeit der ermittelten Beziehung auch über die in die Analyse eingegangenen Werte hinaus, auf.[1] In diesen Fällen empfiehlt es sich, bei größeren Abweichungen von den bisher beobachteten Werten der unabhängigen Variablen eine neue Korrelationsanalyse mit dem veränderten Zahlenmaterial durchzuführen.

Schließlich sei darauf hingewiesen, daß für die statistische Analyse Vergangenheitszahlen, d.h. tatsächlich angefallene Kosten der Vergangenheit (Istkosten) herangezogen werden müssen, die in der Regel nicht frei von zufälligen Schwankungen infolge von Preiseinflüssen, Kostensprüngen, Unwirtschaftlichkeiten oder auch Falschkontierungen sind.[2] Diese Faktoren können den wahren Zusammenhang zwischen abhängiger und unabhängiger Variabler überlagern und verschleiern. Wenn sich solche Zufälligkeiten nicht in genügendem Maß dem Zweck der Rechnung entsprechend ausschalten lassen, muß man damit rechnen, daß unter Umständen eine falsche Schlüsselgröße als die zur Gemeinkostenverrechnung am besten geeignete ermittelt wird, weil sich alle Zufälligkeiten über das in die Analyse eingehende Datenmaterial im Wert des Korrelationskoeffizienten niederschlagen.

Der wichtigste Vorteil der Ermittlung von bestmöglichen Bezugsgrößen mit Hilfe von Korrelationskoeffizienten liegt nach Meinung des Verfassers in der Möglichkeit zur Q u a n -

---

1) Vgl. S. 50 ff.

2) W. KILGER, a.a.O., S. 331 f.

t i f i z i e r u n g der Höhe des Einflusses, den die einzelne untersuchte Größe auf die Kosten hat. Die Festlegung der Schlüsselgrößen nach rein theoretischen Überlegungen - zum Beispiel erscheint es logisch, den Fertigungslohn als Bezugsgröße zu wählen, wenn in der betrachteten Kostenstelle ausschließlich im Akkord gearbeitet wird[1] - kann durch das statistische Verfahren ersetzt oder zumindest ergänzt werden, da dieses einen o b j e k t i v e n Maßstab für die Bestimmung der besten Schlüsselgröße liefert.

4.4.5.2 KOCHs Kritik am Proportionalitätsprinzip und den darauf aufbauenden statistischen Verfahren

Anderer Auffassung als die bisher genannten Autoren bezüglich des Ausgangspunktes für die Bezugsgrößenwahl ist H. KOCH.[2] Er legt dar, daß das Proportionalitätsprinzip überhaupt nicht auf die speziellen Voraussetzungen für die Gemeinkostenschlüsselung abstellt, da es sich auf "eine ganze Skala unterschiedlicher Ausbringungsvolumina bzw. Schlüsselmengen der Abrechnungsperiode und auf die zugehörigen Gemeinkostenfunktionen"[3] bezieht, während die Gemeinkostenschlüsselung nur einen bestimmten Gemeinkostenbetrag und eine bestimmte Gruppe heterogener Leistungseinheiten voraussetzt. Da das Proportionalitätsprinzip dem-

1) Diese rein theoretische Bestimmung der Schlüsselgrößen garantiert keineswegs, die im Sinne des Proportionalitätsprinzips beste Größe gewählt zu haben. Es ist vielmehr möglich, daß die Korrelationskoeffizienten anderer Schlüsselgrößen weit höhere Werte zeigen und sich damit im Sinne unseres Auswahlkriteriums als besser geeignet erweisen.

2) H. KOCH, a.a.O., S. 63 ff.

3) Ebenda. S. 70.

nach unmittelbar keine Aussage über die optimale Schlüsselung trifft, wird der Versuch gemacht, diese Aussage indirekt, nämlich über die Bestimmung des optimalen Beschäftigungsmaßstabes zu gewinnen. KOCH erscheint dieses Vorgehen nicht sinnvoll, da "die Gestalt und Lage der numerischen Gemeinkostenkurve ... als Kriterium für die Wahl des Beschäftigungsmaßstabs unbrauchbar"[1] ist, und zwar aus folgenden Gründen:[2]

1. Die Gemeinkostenkurven sollen mit Hilfe von Buchhaltungs- und Kostenrechnungszahlen empirisch ermittelt werden. Bei dieser Vorgehensweise müssen alle Einflüsse auf die Gemeinkosten mit Ausnahme der betrachteten Bezugsgröße eliminiert werden, da sich lediglich die Veränderungen in dieser Größe und nicht die Veränderungen in anderen Größen (unabhängigen Variablen) auswirken dürfen.[3] Diese Ceteris-paribus-Bedingung wird bei empirischer, d.h. statistischer Ermittlung von Kostenfunktionen oft verletzt, da die zur Ausschaltung anderer Einflüsse erforderlichen Aufbereitungsarbeiten nicht in genügendem Maß durchgeführt werden (können).[4]

2. Da eine Messung des Ausbringungsvolumens durch eine Hilfsgröße nur dann sinnvoll ist, wenn die Auftragszusammensetzung[5] variabel ist, muß die Ceteris-paribus-Klausel in Bezug auf diese Größe eingeschränkt werden. "Wird nun für jeden Beschäftigungsmaßstab unter der Vor-

---

1) H. KOCH, a.a.O., S. 71.

2) Die folgenden Ausführungen sind eng an die Gedankengänge KOCHs angelehnt. Vgl. H.KOCH, a.a.O., S. 70 ff.

3) Ebenda. S. 71 f.; D. SCHNEIDER, a.a.O., S. 682.

4) Vgl. auch S. 75 f.

5) Unter Auftragszusammensetzung wird die Mengenrelation zwischen den heterogenen Produktarten verstanden.

aussetzung einer gegebenen Kapazität die Gemeinkostenkurve in Abhängigkeit von der Periodenmenge dieses Maßstabs bei Varianz der Auftragszusammensetzung aufgestellt, so ergeben sich für jede Periodenmenge so viele verschiedene Gemeinkostenwerte, als unterschiedliche Auftragszusammensetzungen denkbar sind. Demzufolge läßt sich nicht eine eindeutig bestimmte numerische Gemeinkostenkurve, sondern nur ein mehr oder minder breites Gemeinkostenband ermitteln."[1] Aus diesem Grund muß eine Regressionskurve ermittelt werden.

3. Die auf Grund von verschiedenen in Betracht kommenden Bezugsgrößen ermittelten Regressionslinien lassen sich nur bei einheitlichen Voraussetzungen sinnvoll miteinander vergleichen. Diese sind aber nur dann gegeben, wenn die zugrundegelegten Periodenmengen der alternativen Maßstabsgrößen auf derselben Voraussetzung beruhen.[2] Das bedeutet letztlich, daß alle Mengen auf ein bestimmtes Ausbringungsvolumen bezogen werden müssen. In diesem Fall muß man aber beim Vergleich der den alternativen Maßstabsgrößen zugehörigen Gemeinkostenkurven feststellen, daß alle diese Regressionslinien i m P r i n z i p v ö l l i g ü b e r e i n s t i m m e n.[3] Folglich muß auch "der durchschnittliche prozentuale Anteil der proportionalen Gemeinkosten an den Gesamtgemeinkosten bei sämtlichen Maßstabsgrößen gleich sein",[4] so daß sich durch den Vergleich der Korrelationskoeffizienten keine Rangskala für die untersuchten Schlüsselgrößen aufstellen läßt und damit auf diese Weise die optimale Bezugsgröße für die Gemeinkostenverrechnung nicht ermittelt werden kann.

---

1) H. KOCH, a.a.O., S. 72.
2) Ebenda. S. 73.
3) Ebenda.
4) Ebenda.

4.4.5.3 Zusammenfassung

Diese Ausführungen lassen erkennen, daß die Anwendung der Regressions- und Korrelationsanalyse zur Bestimmung der optimalen Schlüsselgröße für die Planung, Kontrolle und Verrechnung von Gemeinkosten zumindest umstritten ist. Die Anhänger des Proportionalitätsprinzips sehen in der Regel in den statistischen Verfahren brauchbare Hilfsmittel, die die Entscheidung zwischen alternativen Bezugsgrößen erleichtern. Sie ermöglichen nämlich - wie gezeigt wurde - über die Berechnung der Korrelationskoeffizienten oder Bestimmtheitsmaße die quantitative Bestimmung der Höhe des Einflusses, den ein Faktor auf die Kosten ausübt.

Demgegenüber lehnt insbesondere KOCH das korrelationsanalytische Verfahren ab.[1] Es zeigt sich, daß in diesem Zusammenhang unter anderem die gleichen Probleme wie bei den statistischen Gesamtkostenuntersuchungen der dreißiger und vierziger Jahre vorliegen, nämlich insbesondere das Problem der Eliminierung von allen Veränderungen außer denen in der untersuchten unabhängigen Variablen (Einhaltung der Ceteris-paribus-Klausel) und der Beseitigung von Zufälligkeiten in den Gemeinkosten, die auf Grund des Vergangenheitscharakters der Beobachtungswerte auftreten.[2] Auch durch umfangreiche und sorgfältige Voruntersuchungen und Aufbereitungsarbeiten ist möglicherweise diesen Forderungen auf Grund der Vielschichtigkeit und engen Verbundenheit betrieblicher Vorgänge nicht immer in vollem Umfang Rechnung zu tragen.

---

1) Vgl. hierzu auch D. SCHNEIDER, Kostentheorie ..., a.a.O., S. 691 ff.

2) Zum Problem der Auswertung von Vergangenheitszahlen vgl. insbesondere Kap. 5., S. 86 ff.

5. Kapitel

ANALYSE VON KOSTEN ZUM ZWECK DER KOSTENPLANUNG

5.1 Einführung

Nach Darstellung der insbesondere für die Perioden- und Stückerfolgsrechnung relevanten Be- und Verrechnungsmethoden im 4. Kapitel wird im folgenden Kapitel auf ein weiteres bedeutendes Aufgabengebiet der Kostenrechnung eingegangen, die Bereitstellung von Zahlen für Planungsrechnungen. Während in den Abschnitten 5.2 und 5.3 die "Zielkalküle" mit der Ermittlung von Plankosten und Kostenbudgets im Vordergrund der Analyse stehen, bilden die "Alternativkalküle" - Wirtschaftlichkeitsrechnungen für kurzfristige Entscheidungen z.B. im Rahmen der Beschäftigungs- und Sortimentsplanung - den Schwerpunkt des folgenden Abschnitts 5.4.[1)]

Es wird gezeigt, daß bei multiplen Zusammenhängen insbesondere die Technik der schrittweisen Regressionsanalyse dem Kostenrechner wertvolle Informationen liefern kann. In einem weiteren Abschnitt wird ausführlich darauf eingegangen, in welcher Weise die statistische Planung der Kosten es erlaubt, die Zuverlässigkeit der Schätzungen durch die Bildung von Konfidenzintervallen wahrscheinlichkeitstheoretisch zu berechnen.

Schließlich soll gezeigt werden, wie die mit Hilfe der Regressionsanalyse ermittelten Kostenabhängigkeiten die Grundlage für ein in mathematische Form gekleidetes Betriebsmodell bilden können.

---

1) Zur Terminologie vgl. S. 6, insbes. aber D. SCHNEIDER, Die wirtschaftliche Nutzungsdauer ..., a.a.O., S. 6 und S. 10.

5.2 Die Ermittlung der Sollkostenkurve (Gemeinkostenplanung) in der Plankostenrechnung

5.2.1 Das Prinzip der flexiblen Gemeinkostenplanung

Die Plankostenrechnung hat in den vergangenen Jahren ständig an Bedeutung gewonnen, da die Unternehmensleitung in immer größerem Maße an der Planung und Kontrolle von Gemeinkosten interessiert sein muß, zum Beispiel um die Idee des "management by exception" verwirklichen zu können.[1] Die Plankostenrechnung ist eine auf die Zukunft gerichtete Form der Kostenrechnung, bei der man mehrere Entwicklungsphasen unterscheiden kann, und die man nach dem Grad ihrer Genauigkeit in starre, flexible, doppelt flexible und voll flexible Rechnung einteilen kann.[2] Im folgenden soll von dem Konzept der flexiblen Plankostenrechnung ausgegangen werden, das als das gebräuchlichste angesehen werden kann, da es die Nachteile des starren Systems vermeidet, ohne wie die doppelt und die voll flexible Plankostenrechnung zu spezielle Anforderungen an den Aufbau des Kostenrechnungssystems zu stellen.

Die wesentliche Aufgabe der flexiblen Plankostenrechnung besteht darin, die Gemeinkosten, nach Kostenarten unterteilt, kostenstellenweise zu planen, wobei diese Planung flexibel, d.h. verschiedenen Beschäftigungsgraden angepaßt, zu erfolgen hat. Die geplanten Kosten werden als Sollgrößen den Abteilungen vorgegeben und die Abweichungen der geplanten von den tatsächlich entstandenen Kosten festgestellt, um der Unternehmensleitung ein Instrument zum Ansporn der Abteilungen und zur Kontrolle der Kosten - oder bei Ausschaltung von Preiseinflüssen durch feste Verrechnungsprei-

---

1) Die Konzeption des "management by exception" setzt eine ständige Beobachtung und Kontrolle des Betriebsgeschehens voraus, um bei Abweichungen vom "Normalen" sofort in den Produktionsablauf eingreifen zu können.

2) Die Definitionen dieser Arten von Plankostenrechnungen findet man in jedem modernen Lehrbuch der Kostenrechnung. Vgl. z.B. W. KILGER, a.a.O., S. 72 f.

se - zur Kontrolle des Faktormengenverbrauchs[1] in die Hand zu geben.[2]

Das bedeutendste Problem im Rahmen dieser Aufgaben ist die eigentliche Planung der Gemeinkosten. Die dafür verwendeten Methoden lassen sich im wesentlichen in vier Gruppen gliedern:[3]

Die Verfahren der ersten Gruppe gehen grundsätzlich von den Istgemeinkosten vergangener Perioden aus und leiten aus ihnen mit Hilfe statistischer Methoden die Sollgemeinkosten ab,[4] weshalb sie auch als historische, empirisch-statistische oder analytische Methoden bezeichnet werden.[5]

Den Verfahren der zweiten Gruppe ist gemeinsam, daß die Sollgemeinkosten mit Hilfe besonderer Kostenuntersuchungen - in der Regel sind dies technische Studien - geplant werden, wobei die Istkosten vergangener Perioden weitestgehend unberücksichtigt bleiben.[6] Sie lassen sich unter dem Begriff "synthetische" Verfahren der Gemeinkostenplanung zusammenfassen.

Als dritte Gruppe lassen sich die Verfahren anführen, bei denen die Aussagen über die Gemeinkosten auf Grund der Ergebnisse von Musteranfertigungen oder Probeläufen getroffen werden, während man als vierte Gruppe die Ermittlung der Sollgemeinkosten anhand von Schätzungen durch Vorarbeiter, Meister oder Spezialisten nennen kann.[7]

---

1) E. KOSIOL, (Hrsg.),: Plankostenrechnung als Instrument moderner Unternehmungsführung. Erhebungen und Studien zur grundsätzlichen Problematik. (Veröffentlichungen des Betriebswirtschaftlichen Instituts für Industrieforschung der Freien Universität Berlin.) Berlin 1956. S. 19.

2) Vgl. z.B. H.-M. SCHÖNFELD: Kostenrechnung, a.a.O., S. 91.; H. VORMBAUM: Kalkulationsarten und Kalkulationsverfahren. Kalkulationslehre. (Betriebswirtschaftliche Studienbücher. Begr. von H. SEISCHAB, Reihe IV: Kaufm. Rechnungswesen.) Stuttgart 1966. S. 42 f.

3) Vgl. K. KÄFER, a.a.O., S. 83.

4) W. KILGER, a.a.O., S. 366.

5) K. KÄFER, a.a.O., S. 83.; H. FUNKE, K. MELLEROWICZ u. H.-G. ABROMEIT, a.a.O., S. 86 f.; W.KILGER, a.a.O.,S. 366.

6) W. KILGER, a.a.O., S. 366.; K. KÄFER, a.a.O., S. 83.

7) K. KÄFER, a.a.O., S. 83.

5.2.2 Die Anwendung der einfachen linearen Regressionsanalyse

Die im Rahmen dieser Arbeit besonders interessierende Frage ist die, inwieweit sich die Gemeinkosten mit den Methoden der ersten Gruppe, den statistischen Verfahren, planen lassen, wobei der Begriff "statistische Verfahren" hier als Regressions- und Korrelationsanalyse konkretisiert werden kann.

Um dieses Verfahren anwenden zu können, müssen für einen längeren Zeitraum Beobachtungswerte für abhängige und unabhängige Variable, d.h. für die Gemeinkostenart und ihre Bezugsgröße, vorliegen. Da die Sollgemeinkostenkurve kostenstellenmäßig ermittelt werden soll, bedeutet das, daß die Istkosten der Vergangenheit nach Kostenstellen und innerhalb der Kostenstellen nach Kostenarten gegliedert aufgezeichnet werden müssen. Die so gewonnenen Beobachtungswerte müssen grundsätzlich vor Durchführung der eigentlichen Regressionsanalyse aufbereitet werden. Die Aufbereitung besteht darin, die Istkosten von allen Einflüssen störender Kostenbestimmungsfaktoren - das sind alle Einflüsse mit Ausnahme desjenigen der betrachteten Bezugsgröße - zu bereinigen, wie zum Beispiel dem Einfluß von Preisschwankungen, Kapazitäts- oder Verfahrensänderungen in der untersuchten Kostenstelle, Fehlkontierungen, Unwirtschaftlichkeiten und sonstige Störungen.[1] Auch hier gilt also die Forderung nach unbedingter Einhaltung der Ceteris-paribus-Bedingung.[2]

KILGER weist mit Recht darauf hin, daß man in diesem Stadium schon nicht mehr von Istkosten der Vergangenheit spre-

1) W. KILGER, a.a.O., S. 366.
2) Vgl. S. 78u. S. 80.

chen kann, sondern daß die beobachteten Kosten durch die verschiedenen Bereinigungen schon mehr oder weniger den Charakter von "geplanten" Kosten angenommen haben.[1)]

Aus diesen "transformierten Istkosten" und der gewählten Bezugsgröße ermittelt man mit Hilfe der Regressionsanalyse die Sollgemeinkosten, wobei überwiegend die einfache l i n e a r e Analyse Anwendung findet, da in der Plankostenrechnung lineare Funktionen vom Typ $y' = a + b\,x$ als ausreichend genau betrachtet werden können.[2)]

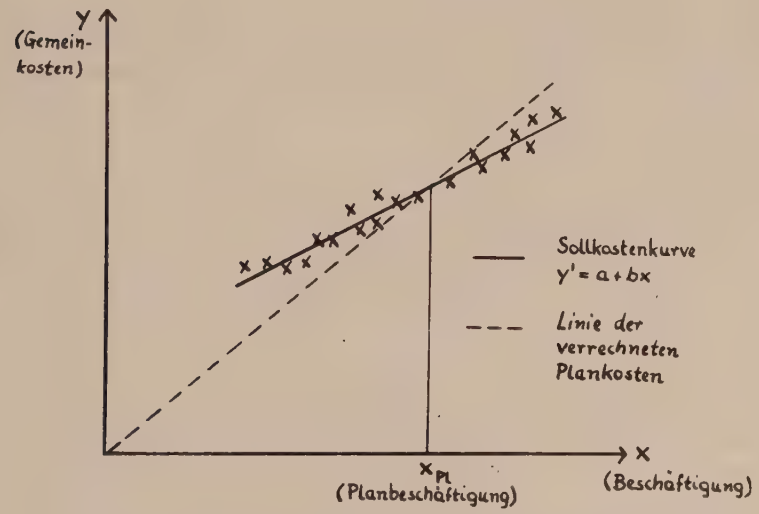


Abb. 5.1: Streupunktdiagramm aus 20 Beobachtungswerten für die Gemeinkosten (y) und die Beschäftigung (x) zur Ermittlung der Sollgemeinkostenkurve.

Die in den Vereinigten Staaten von Amerika relativ häufig anzutreffende Methode der "s c a t t e r c h a r t s"

---

1) W. KILGER, a.a.O., S. 367.

2) Ebenda. S. 370.; K. KÄFER, a.a.O., S. 259.

(Anfertigung eines Streupunktdiagramms und Zeichnung einer "Freihand-Kurve" nach dem Augenmaß durch die Punktwolke) soll hier nicht weiter erörtert werden,[1] da sie im Prinzip lediglich eine ungenaue Form der mathematischen Ermittlung der Koeffizienten durch die Methode der kleinsten Quadrate darstellt und deshalb alle Überlegungen in Bezug auf die Anwendung der Regressionsanalyse bei der Planung von Gemeinkosten in gleicher Weise auch für die Anfertigung von "scatter charts" Gültigkeit haben, sieht man einmal von der Berücksichtigung des entstehenden Arbeitsaufwands für die Durchführung der Analyse ab.

Auf die Ermittlung einer konkreten linearen Regressionsgleichung kann hier verzichtet werden, da die Rechentechnik ausführlich im Abschnitt 2.2 dargestellt worden ist.

5.2.3 Beurteilung des statistischen Verfahrens

5.2.3.1 Vergangenheitszahlen als Ausgangspunkt der Analyse

Wie schneidet nun bei kritischer Beurteilung dieses im vorhergehenden Abschnitt beschriebene statistische Verfahren gegenüber den anderen erwähnten Methoden ab? Um das zu entscheiden, wollen wir in Anlehnung an KILGER fragen,[2] welche Voraussetzungen erfüllt sein müssen, um überhaupt aus Vergangenheitswerten Sollgemeinkosten ableiten zu können, die den Anforderungen der flexiblen Plankostenrechnung genügen.

Es wurde schon kurz darauf hingewiesen, daß die Gemeinkosten als die abhängige Variable nach Kostenstellen und Kosten-

1) Vgl. auch S. 66.

2) Vgl. W. KILGER, a.a.O., S. 372 ff.

arten gegliedert über einen längeren Zeitraum hinweg beobachtet werden müssen, wobei vorauszusetzen ist, daß das während dieser Zeit verwendete Einteilungskriterium auch zukünftig Gültigkeit hat. KILGER ist der Auffassung, daß bei E i n f ü h r u n g einer Plankostenrechnung häufig Verfeinerungen im Kostenarten- und Kostenstellenplan notwendig werden, so daß sich für die statistische Erfassung der Istkosten erhebliche Schwierigkeiten ergeben.[1] Es darf in diesem Zusammenhang aber nicht übersehen werden, daß gerade Vergangenheitszahlen und auf ihnen aufbauende statistische Analysen in vielen Fällen die Einführung einer Gemeinkostenplanung erst e r m ö g l i c h t haben.[2] So wurde in den Vereinigten Staaten in einer Untersuchung der "National Association of Cost Accountants (N.A.C.A.)" festgestellt, daß sich bei der e r s t e n Planung von Stellengemeinkosten von 24 untersuchten Firmen drei auf Vergangenheitszahlen der letzten sechs Monate, zehn auf die aus dem letzten Jahr, acht auf die letzten zwei bis fünf Jahre und drei auf zehn und mehr Jahre gestützt haben.[3] KÄFER ist sogar der Auffassung, daß "weitaus die meisten Betriebe" ihre Gemeinkostenplanung auf der Basis von Vergangenheitszahlen gewonnen haben.[4] Diese Zahlen sagen zwar nichts über die Q u a l i t ä t der mit Hilfe der Regressionsanalyse gewonnenen Schätzwerte aus, lassen jedoch erkennen, daß man in der Praxis - zumindest in den U.S.A. - die Regressionsanalyse als brauchbares Verfahren zur Ermittlung der Sollgemeinkostenkurve ansieht.

---

1) W. KILGER, a.a.O., S. 372.

2) Vgl. K. KÄFER, a.a.O., S. 235.

3) N.A.C.A. - Research Series. How Standard Costs Are Being Used Currently. Standard Costs Research Series No. 11-15. New York (1948) , zitiert bei K. KÄFER, a.a.O., S. 235.

4) K. KÄFER, a.a.O., S. 235.

Die bisher für die Gemeinkosten als der abhängigen Veränderlichen getroffenen Feststellungen gelten in gleicher Weise für die Istbezugsgröße als der unabhängigen Variablen. Auch hier können sich unter Umständen durch Umstellungen in der Kostenstelleneinteilung oder infolge fehlender Aufzeichnungen Schwierigkeiten bei der regressionsanalytischen Ermittlung der Plankosten ergeben.

Sieht man einmal von diesen Problemen ab, so bleiben immer noch die Schwierigkeiten, die bei der Bereinigung der Istkosten von Zufälligkeiten wie Kostenstörungen und Kontierungsfehlern entstehen, sowie Umrechnungsprobleme auf veränderte Preise, Kapazitäten, Arbeitsverfahren und ähnliches mehr.[1] Dies ist ein schwerwiegender Einwand gegen die statistischen Verfahren, weil nur durch eine gründliche und daher umfangreiche Arbeiten umfassende Analyse eine Elimination dieser Störungen möglich ist.[2] Dadurch könnte aber das oft angeführte Argument der besonderen Einfachheit dieser Methode weitgehend entkräftet werden.

Darüber hinaus lassen sich Istkosten vergangener Perioden nur sehr schwer von sog. "Unwirtschaftlichkeiten" bereinigen, da in der Regel - auch wenn ihr Auftreten bekannt ist oder vermutet wird - infolge des Fehlens eines quantitativen Maßstabs keine exakte Aussage über ihren Anteil an der Gesamthöhe der Istkosten getroffen werden kann.[3] Nun ist es aber gerade der Sinn der Aufstellung von Gemeinkostenplänen, durch Abweichungen der tatsächlichen von den ge-

---

1) W. KILGER, a.a.O., S. 372 f.

2) PACK ist sogar der Auffassung, "daß die Eliminierung von Zufälligkeiten nie ganz gelingt". Vgl. L. PACK: Die Elastizität der Kosten. Grundlagen einer entscheidungsorientierten Kostentheorie. (Bd. 1 der Schriften zur theoretischen und angewandten Betriebswirtschaftslehre). Wiesbaden 1966. S. 519.

3) Ebenda.; W. KILGER, a.a.O., S. 373.

planten Kosten Unwirtschaftlichkeiten in den Abteilungen oder Kostenstellen aufzudecken.[1] Dieses Vorhaben muß scheitern, wenn die S o l l k o s t e n selbst aber auf Grund ihrer Ableitung aus vergangenen Istkosten schon Unwirtschaftlichkeiten enthalten. Aus diesem Grund kann das Ziel einer "richtigen" Planung der Sollgemeinkosten nach Auffassung einiger Autoren nur dann erreicht werden, wenn sie losgelöst von Vergangenheitszahlen auf besonderen synthetischen Kostenuntersuchungen aufgebaut werden.[2]

Auf die Tatsache, daß die Regressionsanalyse, und hier insbesondere die multiple Form, auch bei den synthetischen, d.h. bei den auf der Grundlage besonderer technischer Kostenuntersuchungen beruhenden Verfahren der Gemeinkostenplanung Anwendung finden kann, wird im Abschnitt 5.5 eingegangen.[3]

5.2.3.2 Extrapolation der Regressionsgeraden

Ein weiterer Nachteil der statistischen Gemeinkostenplanung kann darin gesehen werden, daß für eine sinnvolle Interpretation der Regressionsparameter - im hier betrachteten Fall linearer Einfachregressionen handelt es sich um die Regressionskonstante a_{yx} und den einfachen Regressionskoeffizienten b_{yx} - von Bedeutung ist, daß die unabhängige Variable im Beobachtungszeitraum relativ starken Schwankungen unterlegen hat, da man nur in diesem Fall einiger-

1) F. WILLE, a.a.O., S. 46.

2) Ebenda. S. 17 u. S. 45 ff.; K. MELLEROWICZ, a.a.O., II, 2. S. 113 u. S. 184 f.; H.-G. PLAUT: Unternehmenssteuerung mit Hilfe der Voll- und Grenz-Plankostenrechnung. "Zeitschrift für Betriebswirtschaft". 31 (1961) 8, S. 460 ff., hier S. 461.; W. KILGER, a.a.O., S. 373 und die dort angegebene Literatur.

3) Vgl. S. 136 ff.

maßen exakte Kostenfunktionen für einen größeren Beschäftigungsbereich ermitteln kann.[1] Auf das Problem der Extrapolation der ermittelten mathematischen Beziehung wurde im Laufe dieser Arbeit schon wiederholt hingewiesen.[2] Dabei wurde festgestellt, daß eine auf der Grundlage eines bestimmten Schwankungsintervalls der Abszissenwerte (Bezugsgröße Beschäftigung) ermittelte Regressionslinie nicht ohne weiteres verlängert, d.h. auf andere niedrigere oder höhere Abszissenwerte ausgedehnt werden darf. Das würde nämlich bedeuten, daß die auf Grund einer ganz bestimmten Kapazitätsauslastung gefundene Beziehung zwischen den Kosten und ihrer Einflußgröße als für s ä m t l i c h e Ausbringungsvolumina[3] (Beschäftigungshöhen) in gleicher Weise gültig unterstellt wird. Hierbei wird nicht berücksichtigt, daß bei Unter- oder Überbeschäftigung die Kosten einen von der angenommenen Linearität abweichenden Verlauf nehmen können, wie die Abbildung 5.2 zeigt.[4]

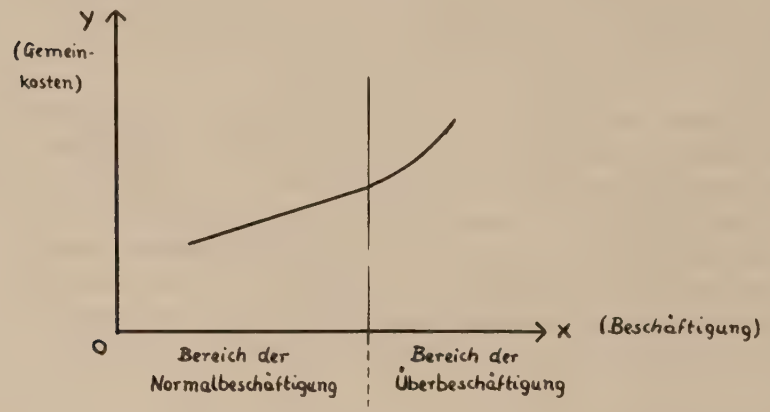


Abb. 5.2: Verlauf der Gemeinkostenkurve bei Normal- und bei Überbeschäftigung, wenn eine intensitätsmäßige Anpassung an Veränderungen der Beschäftigungslage erfolgt.

---

1) Vgl. z.B. W. KILGER, a.a.O., S. 373 f.

2) Vgl. S. 50ff. u. S. 75 f.

3) Unter "sämtlichen Ausbringungsvolumina" sind hier alle von der technischen Kapazität her möglichen Beschäftigungshöhen zu verstehen.

4) Vgl. E. GUTENBERG, a.a.O., (1970) S. 355.

Andererseits läßt sich diese auf den Vergangenheitszahlen basierende Regression gleichwohl in der Zukunft verwenden, wenn man davon ausgehen kann, daß die zukünftigen Istkosten auf Bezugsgrößen basieren, die im gleichen Schwankungsbereich streuen wie die Werte der Vergangenheit. In diesem speziellen Fall taucht das Problem der Extrapolation der Regressionsfunktion nicht auf, und man kann auf Grund der gleichen Ausgangspunkte in Vergangenheit und Zukunft die ermittelten Regreßwerte ohne die oben gemachten Vorbehalte verwenden.

Allerdings darf das zugrundeliegende Schwankungsintervall der unabhängigen Veränderlichen nicht zu klein bemessen sein, da sich sonst die Gefahr einer "Streupunktballung" ergibt,[1] wodurch die Bestimmung einer befriedigenden Regressionsgleichung, d.h. einer Regression mit genügend hohem Korrelationskoeffizienten, unmöglich wird.[2]

5.2.3.3 Bedeutung der Regressionsanalysen in der Praxis

Die im vorigen Abschnitt aufgezählten Kritikpunkte machen deutlich, daß das Verfahren der Regressions- und Korrelationsanalyse bei der Gemeinkostenplanung auf weitgehende Ablehnung stößt und daß von Unwirtschaftlichkeiten und sonstigen Störungskomponenten unbeeinflußte technische oder "ingenieurmäßige" Neuuntersuchungen der Kosten vorzuziehen sind. Berücksichtigt man jedoch die Fülle der Kostenarten, die je Kostenstelle zu planen sind, und den damit verbunde-

1) Vgl. auch W. KILGER, a.a.O., S. 373.

2) Vgl. auch die Abb. 4.4 auf S. 75 und die Ausführungen auf S. 75 f.

nen Arbeitsaufwand, der in der Regel von Spezialisten geleistet werden muß, die nicht in jedem Betrieb zur Verfügung stehen werden, so verliert die - verfahrensmäßig sicherlich wohl fundierte - Forderung nach gedanklicher Neuplanung der Gemeinkosten doch einiges an Nachdruck. Häufig sind Betriebe gar nicht in der Lage, umfangreiche technische Kostenuntersuchungen durchzuführen und müssen sich anderen Methoden zuwenden. Dies geht auch aus einer Untersuchung der "Controllership Foundation" hervor, die in den U.S.A. im Jahre 1958 264 Betriebe nach den bei ihnen verwendeten wichtigsten Verfahren der Gemeinkostenplanung befragte.[1] Dabei ergaben sich die folgenden interessanten Resultate:[2]

67 % : subjektive Schätzungen,
52 % : direkte Beobachtung,
30 % : graphische Methoden,
21 % : Ingenieurstudien,
9 % : Methode der kleinsten Quadrate.

Auch wenn diese Analyse mittlerweile zwölf Jahre zurückliegt und sich die Ergebnisse zugunsten des einen oder anderen Verfahrens verschoben haben mögen, so bleibt die Tatsache, daß der ganz überwiegende Teil aller Betriebe seine Plankosten auf Grund von Vergangenheitszahlen entweder durch subjektive Schätzungen oder durch statistische Verfahren (graphische Methoden bzw. Regressionsanalyse nach der Methode der kleinsten Quadrate) gewonnen hat. Dieses

---

1) B.H. SORD u. G.A. WELSCH: Business Budgeting. A Survey of Management Planning and Control Practices. A Research Study and Report Prepared for Controllership Foundation, Inc., New York 1958. S. 174., zitiert bei K. KÄFER, a.a.O., S. 255.

2) Die Prozentzahlen addieren sich zu mehr als 100 %, da von einzelnen Betrieben mehrere Methoden angewendet wurden.

Ergebnis unterstreicht die Vermutung, daß die meisten Betriebe - wahrscheinlich vornehmlich aus finanziellen Gesichtspunkten - nicht in der Lage sind, groß angelegte synthetische Gemeinkostenuntersuchungen durchzuführen.

Erstaunlich niedrig mit 9 % liegt der Anteil der Betriebe, die sich der Regressionsanalyse in Form der Methode der kleinsten Quadrate bedienten. Dies mag auf den relativ großen Rechenaufwand im Vergleich zu den graphischen Methoden zurückzuführen sein. Berücksichtigt man die heutige Entwicklung und Verbreitung von programmierbaren Tischrechenmaschinen und elektronischen Datenverarbeitungsanlagen sowie die Tatsache, daß viele Hersteller und andere Institutionen Bibliotheksprogramme für Regressions- und Korrelationsanalysen bereithalten,[1] so müßte der Gesichtspunkt der umfangreichen Rechnung weitgehend an Bedeutung verloren haben, zumal die Vermutung nahe liegt, daß Betriebe, die eine Plankostenrechnung einführen wollen oder bereits durchführen, infolge des damit verbundenen Anfalls an Zahlenmaterial auch im Besitz einer Datenverarbeitungsanlage sind.

Im Gegensatz zu der recht seltenen Anwendung der Regressionsanalyse liegt der Anteil der Schätzungen mit 67 % sehr hoch. Solche S c h ä t z u n g e n jedoch sind äußerst problematisch, da sie dem subjektiven Urteil des Schätzenden einen großen Spielraum lassen. Nach Meinung des Verfassers ließe sich dieser hohe Prozentsatz leicht zugunsten des Anteils der Regressionsanalyse an den Verfahren der Gemeinkostenplanung abbauen, die, auch wenn sie aus den angeführten Gründen keine völlig exakten Planungs-

1) Vgl. 9. Anhang I über die rechnerische Durchführung von Regressions- und Korrelationsanalysen, S. 252 ff.

zahlen liefern, doch im Durchschnitt bessere Ergebnisse erwarten lassen als subjektive, nichtstatistische Schätzverfahren.

5.2.3.4 Zusammenfassung

Bei Abwägung der Vor- und Nachteile der Regressionsanalyse bei der Gemeinkostenplanung läßt sich zusammenfassend aussagen, daß sie trotz ihrer offensichtlichen Mängel ein wertvolles H i l f s m i t t e l bei der Planung der Gemeinkosten darstellen kann. Auch theoretisch fundiertere Verfahren können Planungsfehler in gewissem Ausmaß nicht immer vermeiden. Sie können nur durch eine ständige Beobachtung der Istkosten und darauf aufbauenden Angleichungen der geplanten Kosten ausgemerzt werden.[1] Man sollte nach Meinung des Verfassers bei der Beurteilung der Regressionsanalyse als Instrument der Gemeinkostenplanung im Rahmen der flexiblen Plankostenrechnung weniger von der Beachtung aller theoretischen Grundlagen als vielmehr von der praktischen Anwendbarkeit ausgehen. Dann zeigt sich, daß die mit Hilfe der Regressionsanalyse ermittelten Werte in aller Regel eine gute Näherungslösung für die "theoretisch exakten" Plankosten darstellen und damit ausreichenden Ersatz für langwierige und kostspielige Untersuchungen bieten, in besonderem Maße bei Kostenarten, denen im Rahmen der gesamten Kostensumme ein relativ geringes Gewicht zukommt.[2] Bei den Untersuchungen der anderen Kostenarten ist das Verfahren der Regressionsanalyse zumindest als Ergänzung zu anderen Methoden sehr zu empfehlen.

---

1) Vgl. H.-G. PLAUT: Die Grenz-Plankostenrechnung. "Zeitschrift für Betriebswirtschaft". 23 (1953) 6, S. 347 ff., hier S. 349.

2) Anderer Meinung ist W. KILGER, a.a.O., S. 372 f.

Darüber hinaus bietet sich bei Anwendung der Regressionsanalyse für den Kostenrechner der Vorteil, für die Regressionsparameter und die Schätzwerte der Gemeinkosten Konfidenzintervalle bilden und auf diese Weise eine Art "Kontrollkarte der Gemeinkosten" aufstellen zu können, die die Grundlage für die Konzeption des "management by exception" bilden kann.[1)]

5.3 Aufstellung von Kostenbudgets

5.3.1 Begriff der Budgetkosten im Vergleich zu Plankosten

Während die Plankosten vorwiegend dazu dienten, durch Vergleich mit den tatsächlich entstandenen Istkosten Unwirtschaftlichkeiten beim Kostengüterverbrauch aufzudecken, soll mit Hilfe von Budgetkosten gezeigt werden, "wieweit sich das tatsächliche Betriebsgeschehen im Rahmen des Geplanten hält oder wo und in welchem Umfang Veränderungen und Störungen auftreten, die entsprechende Maßnahmen verlangen".[2)] Geht man von diesen Begriffsinhalten aus, so läßt sich die Plankostenrechnung auch als Vergleichsrechnung zum Zweck der Kontrolle der Betriebsgebarung, die Budgetkostenrechnung dagegen als V o r s c h a u r e c h n u n g z u D i s p o s i t i o n s z w e c k e n definieren.[3)] Damit ist der Begriffsinhalt der Budgetkosten ein anderer als der der Plankosten. Während diese sich durch die Begriffe "Verbrauchsnormen" oder "angestrebte Sollkosten" charakterisieren lassen, liegen den Budgetkosten "V e r b r a u c h s e r w a r t u n g e n" oder "z u k ü n f t i g e I s t k o s t e n" zugrunde.[4)]

---

1) Diese Technik wird ausführlich im Rahmen der Kostenkontrolle im Kap. 6 dargestellt. Vgl. S. 168 ff.

2) K. MELLEROWICZ, a.a.O., II, 2. S. 252.

3) In Anlehnung an K. MELLEROWICZ, a.a.O., II, 2. S. 154.

4) Ebenda.

Im Abschnitt 5.2 wurde dargelegt, daß einer der entscheidenden Kritikpunkte an den statistischen Verfahren der Gemeinkostenplanung auf der Tatsache beruht, daß sie sich an tatsächlich entstandenen Kosten der Vergangenheit orientieren. Dieser Gesichtspunkt kann bei der statistischen Ermittlung von Budgetkosten nicht mehr in gleichem Maße negativ bewertet werden, da sie t a t s ä c h l i c h e n t s t e h e n d e zukünftige Kosten - und keine Verbrauchsnormen - repräsentieren sollen, die in der Regel recht gut auf Grund von Istkosten der Vergangenheit ermittelt werden können.[1)]

Allerdings haben trotz der hier bewußt vorgenommenen scharfen Trennung von Budget- und Plankosten beide zahlreiche Beziehungen zueinander, so daß in Betrieben mit ausgebauter Plankostenrechnung die Plankosten durchaus unter Berücksichtigung von Zu- und Abschlägen für die Ermittlung der Budgetkosten zu Hilfe genommen werden können.[2)]

Legt man einer Vorschaurechnung Vergangenheitszahlen zugrunde, dann sind allerdings zukünftige Erwartungen zu berücksichtigen, so daß in der Regel die Istkosten der Vergangenheit nicht ohne Korrekturen übernommen werden dürfen.[3)] Diese Korrekturen oder Bereinigungen der Istkosten beziehen sich aber nicht wie bei der Ermittlung der Plankosten auf die Ausschaltung von Unwirtschaftlichkeiten,

---

1) K. MELLEROWICZ, a.a.O., II, 2. S. 154.

2) Ebenda. S. 159 f.

3) Vgl. z.B. R.E. JENSEN: A Multiple Regression Model for Cost Control - Assumptions and Limitations. "The Accounting Review". 42 (1967) 2, S. 265 ff., hier S. 268, wo es heißt:"... hängt die Aussagefähigkeit des Schätzwertes für die Kosten von der Stabilität des Systems von Periode zu Periode ab, da die Werte der Variablen auf laufendem [gegenwärtigem] und nicht auf zukünftigem Datenmaterial beruhen." (Sinngemäße Übersetzung des Verfassers).

sondern lediglich auf die Berücksichtigung von zukünftig erwarteten Veränderungen, die sich naturgemäß nicht in den in der Vergangenheit beobachteten Kosten niedergeschlagen haben können.[1)]

Unter Einbeziehung dieser Erwartungen lassen sich demnach Budgetkosten relativ gut aus beobachteten Kosten der letzten Abrechnungsperioden ermitteln, wobei die Regressions- und Korrelationsanalyse wertvolle Hilfe leisten kann.[2)]

5.3.2 Die Bedeutung kurzfristiger Vorschaurechnungen

Nach MELLEROWICZ finden Budgetkosten vorwiegend im Beschaffungs-, Verwaltungs- und Vertriebsbereich Anwendung, während man die Plankostenrechnung hauptsächlich im Fertigungsbereich antrifft.[3)] Häufig werden auch für Tochtergesellschaften von großen Unternehmen Budgets aufgestellt,[4)] die nicht oder nur in geringem Umfang überschritten werden dürfen, um eine enge Kooperation mit der Muttergesellschaft zu garantieren. Auf diese Weise läßt sich auch bei dezentralisierter Organisation eine verhältnismäßig exakte gesamtbetriebliche Planung realisieren.

Aber auch im Fertigungsbereich kann die Ermittlung von Budgetkosten zum Zweck der Kostenplanung wertvolle Dienste

---

1) Vgl. dazu die Vorteile, die die schrittweise Regressionsanalyse bietet. Vgl. S. 99 ff.

2) Vgl. in diesem Zusammenhang auch R.B. TROXEL: Variable Budgets Through Correlation Analysis. A Simplified Approach. "National Association of Accountants (N.A.A.)-Bulletin". New York (1965). S. 48 ff.

3) K. MELLEROWICZ, a.a.O., II, 2. S. 154.

4) Vgl. R.E. JENSEN, a.a.O., S. 268.; G.J. BENSTON: Multiple Regression Analysis of Cost Behavior. "The Accounting Review". 41 (1966) 4, S. 657 ff., hier S. 658 f.

leisten. Es ist eine bekannte Tatsache, daß durch die herkömmlichen Formen des industriellen Rechnungswesens - Istkostenrechnung, flexible Plankostenrechnung, Grenzplankostenrechnung oder ähnliche Formen der Kostenrechnung - die Zwecke der Dokumentation und Wirtschaftlichkeitskontrolle sowie die Bewertung der Bestände (für Bilanzansätze) in ausreichender Weise erfüllt werden. Dagegen sind kurzfristige Vorschaurechnungen und zuverlässige Planungen mit den genannten Kostenrechnungsverfahren nur schwer oder gar nicht möglich.

Für kurzfristige Vorschaurechnungen müssen dem Unternehmer nämlich Zahlen zur Verfügung stehen, die es erlauben, Entscheidungen über alternative Produktionsverfahren, Stillegung von Maschinen und Aggregaten, Sortimentszusammensetzungen und ähnliches mehr k u r z f r i s t i g zu treffen,[1] d.h. bereits dann, wenn seine Entscheidung noch Einfluß auf die Entwicklung nehmen kann. Zur Ermittlung des Datenmaterials, das die Grundlage für solche Entscheidungen bildet, sind insbesondere Kenntnisse über Abhängigkeiten oder Gesetzmäßigkeiten zwischen Kosten oder Verbrauchsmengen und den ihre Höhe beeinflussenden Faktoren erforderlich.

Aus diesem Grund werden seit einiger Zeit insbesondere in der Eisen- und Stahlindustrie Versuche unternommen, Abhängigkeiten zwischen einzelnen Kostenarten oder Verbrauchsmengen und ihren Einflußfaktoren mit Hilfe der einfachen oder multiplen Regressionsanalyse in mathematischer Form auszudrücken. Das Ziel dieser Untersuchungen besteht darin, für eine Kostenstelle, eine Betriebsabteilung oder sogar einen ganzen Betrieb ein in mathematischer Weise formuliertes geschlossenes Rechnungssystem zu entwickeln, das insbe-

---

1) Vgl. hierzu auch den Abschnitt 5.4, S. 118 ff.

sondere für Zwecke der kurzfristigen Planung, aber auch der Dokumentation und Kontrolle, in Industrieunternehmen mit Massen- und Sortenfertigung eingesetzt werden kann.[1)]

5.3.3 Die schrittweise Regressionsanalyse

Um zu zuverlässigen empirischen Schätzwerten für die Abhängigkeiten zwischen Kosten oder Verbrauchsmengen und ihren Einflußfaktoren zu gelangen, ist in aller Regel eine multiple Regressionsanalyse mit mehreren unabhängigen Variablen (Einflußgrößen) erforderlich. Die Kosten werden nämlich von vielen Größen beeinflußt oder verursacht, die sich nicht zu einer einzigen unabhängigen Veränderlichen zusammenfassen lassen, sondern jede für sich einen getrennten, andersartigen Effekt ausüben. Zur Durchführung dieser multiplen Regressionsanalysen bietet sich ein Anfang der sechziger Jahre entwickeltes Verfahren an, das bei der Ermittlung der Koeffizienten etwas anders vorgeht als es bei der bisher beschriebenen Methode der kleinsten Quadrate geschieht. Dieses Verfahren ist unter dem Namen "schrittweise Regressionsanalyse" bekannt geworden.[2)]

5.3.3.1 Darstellung des Verfahrens

Auch die schrittweise Regressionsanalyse baut wie das

---

1) Vgl. u.a. G. LASSMANN: Die Kosten- und Erlösrechnung als Instrument der Planung und Kontrolle in Industriebetrieben. Betriebswirtschaftliches Institut der Eisenhüttenindustrie. Düsseldorf 1968. S. 72 ff. Vgl. auch S. 161 ff. dieser Arbeit.

2) Vgl. z.B. A.S. GOLDBERGER u. D.B. JOCHEMS: Note on Stepwise Least Squares. "Journal of the American Statistical Association". 56 (1961) 293, S. 105 ff.; N.R. DRAPER u. H. SMITH: Applied Regression Analysis. Third printing. New York/London/Sydney 1967. S. 171 ff..

bisher beschriebene Verfahren[1] auf der Methode der kleinsten Quadrate auf. Man bildet zunächst - wie das auch bei der "herkömmlichen" Regressionsanalyse geschieht - einen Gleichungsansatz mit allen in Frage kommenden unabhängigen Variablen, führt die Rechnung dann aber s c h r i t t - w e i s e durch. Das bedeutet, daß man aus der möglichen Vielzahl von Variablen zunächst e i n e Variable herausucht und zwar diejenige, die den größten Einfluß auf die Zielgröße ausübt. Das ist nach dem Kriterium der kleinsten Quadrate d i e unabhängige Variable, "die die Standardabweichung zwischen den berechneten und den beobachteten Werten der Zielgröße am stärksten minimiert".[2] Löst man das Gleichungssystem nach dieser Einflußgröße auf, so erhält man als erstes Ergebnis der Rechnung eine Regressionsgleichung zwischen der Zielgröße (abhängige Variable) und der - im Sinne der Methode der kleinsten Quadrate - wichtigsten Einflußgröße. Damit ist der erste Schritt der Rechnung beendet.

Im zweiten Schritt sucht man auf die gleiche Weise die unabhängige Variable mit dem nächsthöheren Einfluß auf die Zielgröße, d.h. die Größe, die in diesem Schritt die Summe der Abweichungsquadrate zwischen den beobachteten und den berechneten Werten der abhängigen Variablen am stärksten verkleinert.[3] Löst man das Gleichungssystem nach dieser Einflußgröße auf, so erhält man eine Regressionsgleichung mit z w e i unabhängigen Variablen. Diese beiden Veränderlichen sind die "wichtigsten", d.h. diejenigen mit dem größten Einfluß auf die Zielgröße.

---

1) Vgl. S. 13 ff.

2) G. MATT: Die schrittweise Regressionsanalyse ..., a.a.O., S. 256. (Diese Schrift ist auch erschienen als Sonderdruck der IBM Deutschland, IBM Form 78132.)

3) Ebenda. S. 257.

Diese Berechnungsweise wird Schritt für Schritt fortgeführt, bis alle wichtigen, d.h. alle signifikanten Einflußgrößen in der Regression enthalten sind. Als signifikant werden alle diejenigen unabhängigen Variablen bezeichnet, die eine vorher festgelegte statistische Sicherheit aufweisen,[1] wobei die Entscheidung über Signifikanz oder Nichtsignifikanz einer Einflußgröße - und damit über die Aufnahme in die Vorhersagegleichung - anhand der Berechnung von t-Werten oder F-Werten für diese Variablen und ihrem Vergleich mit dem vorher festgelegten Testwert erfolgt.[2] So kann man zum Beispiel bestimmen, daß ein t-Wert von 2,0 von keiner Einflußgröße unterschritten werden darf, was einem Konfidenzniveau von 95 % (Signifikanzniveau von 5 %) bei 60 oder mehr Freiheitsgraden entspricht.[3] Das bedeutet, daß nur die Regressionskoeffizienten in die Regressionsgleichung aufgenommen werden, für die sich ein Wert von $t \geq 2{,}0$ errechnet.[4] Verliert während der Rechnungsschritte eine schon in die Vorhersagegleichung aufge-

---

1) Vgl. G. MATT: Bestimmung statistisch gesicherter Koeffizienten bei der exponentiellen Ausgleichung (Exponential Smoothing). "Unternehmensforschung - Operations Research" 10 (1966) 1, S. 15 ff., hier S. 23.

2) Zur Ermittlung von t- oder F-Werten zur Durchführung von Signifikanzprüfungen vgl. insbesondere die Ausführungen im Abschnitt 5.6, S. 141 ff. und im Abschnitt 6.2, S. 168 ff. Die F-Verteilung wird oft aus rechentechnischen Gründen verwendet, um der Vorzeichenproblematik bei den t-Werten zu entgehen.

3) Vgl. im Anhang II die Tabelle der Studentverteilung, Tabelle 4* und die Tabelle der Standardnormalverteilung, Tabelle 2*, S. 285 ff.

4) Vgl. in diesem Zusammenhang auch die Ausführungen im Anhang I: Die rechnerische Durchführung von Regressions- und Korrelationsanalysen, wo auf die Bedeutung von Tests für die partiellen Regressionskoeffizienten ausführlich eingegangen wird, insbesondere bei der Darstellung des IBM-Programms "STEPR", S. 265 f. und des Programms "MLRP" der Bull General Electric, S. 271 ff.

nommene Einflußgröße diese statistische Sicherheit - durch Korrelation mit einer zu einem späteren Zeitpunkt in die Gleichung aufgenommenen Variablen[1] - so wird die Ausgangsvariable aus der Regressionsgleichung als nicht mehr signifikant wieder eliminiert.[2] Sie wird wieder bei den Größen aufgenommen, die noch nicht in der Gleichung enthalten sind. Dadurch besteht die Möglichkeit, sie in einem späteren Schritt doch noch in die endgültige Regressionsgleichung aufzunehmen, falls nämlich wiederum andere, bis dahin in der Gleichung enthaltene - und mit ihr korrelierte - Größen durch Korrelation mit neuen unabhängigen Variablen ausscheiden müssen.[3]

In mathematischer Form lassen sich die Grundzüge der schrittweisen Regressionsanalyse folgendermaßen darstellen.[4] In Abweichung von der bisher gewählten Symbolik wird die Regressionskonstante nicht mit "a" sondern mit dem Symbol "b_o" bezeichnet, um die Schreibweise der folgenden Gleichungen zu erleichtern. Den Ausgangspunkt der schrittweisen Analyse bildet die Gleichung

$$y' = b_o + b_1x_1 + b_2x_2 + \dots + b_mx_m = \sum_{k=0}^{m} b_kx_k \,, \quad (5.1)$$

---

1) Zum Problem der "Multikollinearität" vgl. den Abschnitt 5.3.4, S. 111ff.

2) G. MATT: Die schrittweise Regressionsanalyse ..., a.a.O., S. 256.; A. RALSTON u. H.S. WILF: Mathematische Methoden für Digitalrechner. Genehmigte Übersetzung der im Verlag John Wiley & Sons, Inc., New York, unter dem Titel "Mathematical Methods for Digital Computers" erschienenen englischsprachigen Ausgabe. München und Wien 1967. S. 346.; Vgl. auch 9. Anhang I, S. 252 ff.

3) Zur praktischen Berechnung der Regressionsgleichung vgl. z.B. A. LINDER: Statistische Methoden. Für Naturwissenschaftler, Mediziner und Ingenieure. (Bd. 3 der mathematischen Reihe der Lehrbücher und Monographien aus dem Gebiete der exakten Wissenschaften). 4., unveränd. Aufl. Basel und Stuttgart 1964. S. 196 ff., wo ein Rechenschema mit Kontrollmöglichkeiten entwickelt wird. Vgl. aber auch den Anhang I, insbesondere S. 265 ff. u. S. 271 f.

4) Die Darstellung wurde unter Verwendung anderer Symbolik übernommen von G. MATT: Die schrittweise Regressionsanalyse ..., a.a.O., S. 257 ff.

in der y^i den Schätzwert für die Kosten oder Verbrauchsmengen und die x_k die verschiedenen Einflußgrößen (Produktionsmengen, Arbeitsstunden oder ähnliche Größen) darstellen.

Aus der partiellen Differentiation der Bedingung für ein Minimum der Summe der Abweichungsquadrate erhält man (m + 1) Gleichungen, nämlich eine für das absolute Glied b_o und je eine Gleichung für jeden Regressionskoeffizienten b_k.[1] Quadriert man in einem weiteren Schritt die Bedingungsgleichung für das Minimum

$$S = \sum_{i=1}^{n} (y_i - y_i')^2 = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \sum_{k=0}^{m} b_k x_{ik})^2 \quad (5.2)$$

$$= \text{Min.}$$

unter Berücksichtigung der Gleichungen für die partiellen Ableitungen aus, so erhält man ein lineares Gleichungssystem mit (m + 2) Gleichungen:[2]

$$\begin{aligned}
&b_o n + b_1[x_1] + \cdots\cdots\cdots\cdots + b_m[x_m] = [y] \\
&\cdot \\
&\cdot \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (5.3) \\
&\cdot \\
&b_o[x_m] + b_1[x_1 x_m] + \cdots\cdots\cdots + b_m[x_m^2] = [y x_m] \\
&b_o[y] + b_1[x_1 y] + \cdots\cdots + b_m[x_m y] + [e^2] = [y^2]
\end{aligned}$$

---

1) Vgl. auch die Darstellung der einfachen Regression auf S. 13 ff.

2) Es wird im folgenden die abgekürzte GAUSSsche Schreibweise gewählt. Dabei gilt:

$$\sum_{i=1}^{n} x_{ik} x_{ij} = [x_k x_j].$$

Schreibt man nun für $\left[x_j x_k\right] = a_{jk}$ und fügt außerdem noch eine Einheitsmatrix hinzu, die der Berechnung der inversen Matrix[1] und über diese der Ermittlung der Standardabweichungen der Regressionskoeffizienten und der Ausschaltung nichtsignifikanter Einflußgrößen dient,[2] so kann man schreiben:

$$\begin{array}{llll} a_{oo} & a_{o1} \cdots\cdots\cdots a_{o,m} & a_{o,m+1} & 1 \cdots 0\ 0 \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ a_{m,o} & a_{m,1} \cdots\cdots a_{m,m} & a_{m,m+1} & 0 \cdots 1\ 0 \\ a_{m+1,o} & a_{m+1,1} \cdots\cdots a_{m+1,m} & a_{m+1,m+1} & 0 \cdots 0\ 1 \end{array} \qquad (5.4)$$

Diese Matrix entwickelt man schrittweise mit Hilfe des GAUSSschen Algorithmus nach den Diagonalgliedern. Bei jedem Schritt wird das Diagonalglied ausgewählt, das $a_{m+1,m+1} = \left[e^2\right]$ am stärksten verkleinert. Mit anderen Worten, man sucht schrittweise jeweils die unabhängige Variable heraus, die den größten Einfluß auf die Zielgröße besitzt, da sie die Summe der Abweichungsquadrate zwischen berechneten und beobachteten Werten der Zielgröße am stärksten verringert.[3]

Diese kurze Darstellung der Grundlagen der schrittweisen Regressionsanalyse kann im Rahmen dieser Arbeit genügen, da für den Kostenrechner weniger die mathematische Konzeption als vielmehr die Vorteile dieses Verfahrens bei der

---

1) Zur Matrizenrechnung vgl. z.B. A. KLINGST: Einführung in die Matrizenrechnung. In: Anwendungen der Matrizenrechnung auf wirtschaftliche und statistische Probleme. 3. Aufl. Würzburg und Wien 1966. S. 9 ff.

2) Vgl. hierzu auch 9. Anhang I über die rechnerische Durchführung von Regressions- und Korrelationsanalysen, insbesondere S. 271 f.

3) Vgl. G. MATT: Die schrittweise Regressionsanalyse ..., a.a.O., S. 257 f.

Anwendung auf seine Probleme von Interesse sein dürften. Hierauf wird in den folgenden Abschnitten ausführlich eingegangen.[1)]

5.3.3.2 Vorteile gegenüber der herkömmlichen Methode

Die schrittweise Regressionsanalyse weist besonders in den Fällen, in denen unabhängige Variable untereinander korreliert sind, einige bedeutende Vorteile gegenüber der herkömmlichen Berechnungsweise nach der Methode der kleinsten Quadrate auf.

In den Gleichungsansatz können beliebig viele Variable, die sogar transformiert oder miteinander kombiniert sein können, aufgenommen werden. Trotzdem ist über die Bestimmung der Testwerte in jedem Schritt der Rechnung und eventuelle Ausschaltung von Einflußgrößen gewährleistet, daß die endgültige Lösungsgleichung nur noch solche unabhängigen Veränderlichen enthält, die einen statistisch gesicherten Einfluß auf die abhängige Variable haben.[2)] Zu demselben Ergebnis kann man durch die "normale" Berechnungsweise der Regressionskoeffizienten nur dann gelangen, wenn man diese "statistisch gesicherten" Größen sofort in den Glei-

---

1) Die Beschränkung auf die knappe Darstellung der mathematischen Grundlagen des Modells der schrittweisen Regressionsanalyse erscheint dem Verfasser auch sinnvoll, da diese Form der Regressionsanalyse einen erheblichen Rechenaufwand erfordert, der nur durch den Einsatz elektronischer Datenverarbeitungsanlagen in ökonomischer Weise bewältigt werden kann. Da für diese Probleme ausgezeichnete "software" für EDV-Anlagen zur Verfügung steht, kann der Kostenrechner sich ganz auf die betriebswirtschaftliche Interpretation der von der Maschine gelieferten Ergebnisse konzentrieren, ohne die Durchführung der Rechnung selbst in allen Einzelheiten beherrschen zu müssen.

2) G. MATT: Die schrittweise Regressionsanalyse ..., a.a.O., S. 256.; G. MATT: Bestimmung ..., a.a.O., S. 26.

chungsansatz aufnimmt. Das jedoch wird in der Regel äußerst schwierig sein, da man gerade durch die Regressions- und Korrelationsanalyse erst herausfinden will, welche Größen die zu untersuchende Zielgröße am stärksten beeinflussen, denn häufig fehlen frühere Untersuchungen oder andere Kenntnisse, um sofort den bestmöglichen Gleichungsansatz wählen zu können.

Ein zweiter Vorteil der schrittweisen Regressionsanalyse liegt in der Tatsache, daß neben der endgültigen Gleichung der mehrfachen Regression bei jedem Schritt eine weitere "Zwischenregressionsgleichung" entsteht.[1] Auch diese Gleichungen, die - berücksichtigt man die Zahl der in ihnen enthaltenen unabhängigen Variablen - jeweils die "besten" im Sinne der Methode der kleinsten Quadrate sind[2] - können für einen Betrieb von Interesse sein. So sind häufig mehrere verschiedene Betriebsabteilungen an den Regressionsergebnissen interessiert, jede auf Grund ihres speziellen Aufgabengebietes. Daraus können verschiedenartige Anforderungen an die Regressionsgleichung resultieren. So bevorzugt eine Stelle eine Regressionsgleichung mit möglichst wenigen Einflußfaktoren, ist aber dafür bereit, einen größeren statistischen Fehler in Kauf zu nehmen.[3] Eine andere Stelle interessiert sich für mehrere Gleichungen, um anhand von zukünftigen Beobachtungswerten diese Gleichungen auf ihre Zuverlässigkeit in Bezug auf Schätzungen für die

---

1) G. MATT: Die schrittweise Regressionsanalyse ..., a.a.O., S. 256 f.; A. RALSTON u. H.S. WILF, a.a.O., S. 346.

2) Die Aufnahme der unabhängigen Variablen erfolgte nämlich in der Reihenfolge der jeweils größtmöglichen Reduktion in der Summe der Abweichungsquadrate.

3) Der prozentuale Anteil der Veränderungen in der Zielgröße, der durch die Veränderungen in den unabhängigen Variablen erklärt wird, ist nicht so hoch wie bei der endgültigen Regressionsgleichung (kleinerer Korrelationskoeffizient).

Zukunft zu testen. Häufig hat sich gezeigt, daß nicht immer die Gleichung in der Zukunft die besten Schätzwerte liefert, die die Vergangenheitswerte am besten repräsentiert, da sich durch zukünftige Entwicklungen (wie Änderungen eines Produktionsverfahrens) eine bisher relativ unwesentliche Einflußgröße, die auf Grund ihres niedrigen t-Wertes nicht in die Vorhersagegleichung aufgenommen wurde, für die Zukunft als sehr bedeutend erweisen kann.[1)]

Zur Durchführung der Rechnung auf elektronischen Datenverarbeitungsanlagen - nur eine solche Berechnungsweise kann bei der sehr umfangreichen Rechenarbeit überhaupt in Erwägung gezogen werden - stehen für die schrittweise Regressionsanalyse zahlreiche Bibliotheksprogramme zur Verfügung, so daß die Durchführung der Analyse selbst in kurzer Zeit zu bewältigen ist.[2)]

5.3.3.3 Die Anwendung der schrittweisen Regressionsanalyse

Ein interessantes Beispiel für die Anwendung der schrittweisen Regressionsanalyse im Rahmen der industriellen Fertigung hat MATT gegeben, der in einem Hüttenwerk den Koksverbrauch bei der Thomasroheisenerzeugung für Zwecke der kurzfristigen Vorschaurechnung zu planen versuchte.[3)] Bei zwei nacheinander durchgeführten Untersuchungen wurden in der ersten 14 und in der zweiten 15 unabhängige Variable

---

1) Vgl. G. MATT: Die schrittweise Regressionsanalyse ..., a.a.O., S. 262.

2) Vgl. A. RALSTON u. H.S. WILF, a.a.O., S. 344., insbesondere aber 9. Anhang I, S. 252 ff.

3) G. MATT: Die schrittweise Regressionsanalyse ..., a.a.O., S. 261 ff.

in den Gleichungsansatz aufgenommen, wobei der Berechnung 36 bzw. 31 Monatszahlen der Einflußgrößen und der Zielgröße (Stichprobenumfang) zugrundelagen.

Beim Vergleich der mittels statistischer Analyse gewonnenen Schätzwerte mit dem in der Folgezeit tatsächlich gemessenen Verbrauch zeigte sich, daß eine gute Übereinstimmung vorlag. Dabei erwies sich die für die schrittweise Regressionsanalyse charakteristische Vorgehensweise mit der Ermittlung von mehreren "Zwischen-Regressionsgleichungen" als besonders vorteilhaft. Da man durch die sukzessive Aufnahme weiterer Einflußgrößen in die Ausgangsgleichung mehrere Regressionen mit etwa der gleichen Standardabweichung und etwa der gleichen statistischen Sicherheit der Regressionskoeffizienten erhält, kann man in den der Analyse folgenden Monaten die verschiedenen Gleichungen testen und diejenige herausfinden, die die z u k ü n f t i g e n Veränderungen am besten erfaßt. Dieser Gesichtspunkt ist für die Anwendung von Regressionsanalysen zur Planung von Kosten oder Verbrauchsmengen von großer Bedeutung. Die schrittweise Analyse bietet die Möglichkeit, aus V e r g a n g e n h e i t s z a h l e n P l a n u n g s w e r t e abzuleiten, indem durch laufende Anpassung der Vorhersagegleichung an in der Zukunft beobachtete Werte eine ständige Modifizierung der "besten" Gleichung erfolgt. Das ist in diesem Ausmaß nur möglich, wenn ständig verschiedene Regressionsgleichungen zur Auswahl zur Verfügung stehen, wie sie die schrittweise Regressionsanalyse bereitstellt.

Bei der von MATT durchgeführten Untersuchung[1)] zeigte sich zum Beispiel, daß die Regressionsgleichung mit dem höchsten

---

1) Es wurden 14 unabhängige Variable in die Rechnung einbezogen, der Stichprobenumfang betrug n = 36.

statistischen Konfidenzniveau,[1] die fünf unabhängige Variable enthielt, für Prognosezwecke nicht geeignet war. In ihr war nämlich der Sinter-Einsatz - der in den folgenden Monaten stark erhöht wurde - auf Grund eines geringeren Konfidenzniveaus seines Regressionskoeffizienten - nicht enthalten. Die beste Übereinstimmung mit den zukünftig beobachteten Verbrauchsmengen ergab dagegen eine Gleichung mit sieben unabhängigen Variablen - unter Einbeziehung des Sinter-Einsatzes - wenn auch zwei Einflußgrößen einen Wert für $t < 2{,}0$ aufwiesen. Bei einem kleinsten t-Wert von 1,59 - das entspricht immer noch einem Konfidenzniveau von 88,8 % - kann man aber die statistische Sicherheit der Koeffizienten immer noch als genügend ansehen.[2]

Auf die Durchführung eines eigenen Zahlenbeispiels für die schrittweise Regressionsanalyse soll hier verzichtet werden, da die für die Planung von Kosten entscheidenden Feststellungen anhand des zitierten Beispiels von MATT getroffen werden konnten. Ausführliche Rechenbeispiele findet der an den einzelnen Schritten der Rechnung und der Aufnahme, Ausschaltung und Wiederaufnahme von Einflußgrößen interessierte Leser in den Veröffentlichungen der Herstellerfirmen für elektronische Datenverarbeitungsanlagen.[3]

---

1) Der kleinste t-Wert für die fünf Regressionskoeffizienten betrug $t = 2{,}95$, so daß alle Koeffizienten eine statistische Sicherheit von über 99 % aufwiesen. Vgl. zu Fragen der statistischen Sicherheit insbesondere den Abschnitt 5.6., S. 141 ff.

2) Vgl. auch G. MATT: Die schrittweise Regressionsanalyse.., a.a.O., S. 262.

3) Vgl. z.B.: IBM Application Program. System/360 Scientific Subroutine Package. (360 A - CM - 03 x). Version III. Programmer's Manual. Fourth edition. Hrsg. von International Business Machines Corporation 1966, 1967, 1968. S. 415 f.; Multiple Linear Regression Program. GE - 400 Series. Application Manual. Hrsg. von Bull General Electric. O.O., Oktober 1965, durchges. Aufl. September 1969. S. 63 ff. Vgl. auch den Anhang I zur rechnerischen Durchführung von Regressions- und Korrelationsanalysen, insbesondere S. 265 ff. u. S. 271 ff.

Die schrittweise Regressionsanalyse mit ihrer Ermittlung von zahlreichen unterschiedlichen Gleichungen - unterschiedlich in Bezug auf die Berücksichtigung von unabhängigen Variablen und die statistische Sicherheit der Koeffizienten - versetzt den Kostenplaner damit in die Lage, Strukturänderungen in der Zukunft durch die Auswahl der geeigneten Regressionsgleichung Rechnung zu tragen. Am besten geeignet ist immer die Gleichung, durch die die Veränderung über den Einbau der betreffenden Einflußgröße am besten wiedergegeben wird. Allerdings genügt es nicht, lediglich Veränderungen einer oder weniger Folgeperioden zu berücksichtigen. Um eine ausreichende statistische Sicherheit zu erhalten, müssen genügend viele Beobachtungsperioden ausgewertet werden.[1)]

Auf der gleichen Ebene - vom Standpunkt dieser Arbeit aus betrachtet - wie die Aufstellung von Kostenbudgets liegt die Ermittlung von sog. "N o r m a l k o s t e n". Als Normalkosten bezeichnet man üblicherweise die Kosten, die zwar auf Vergangenheitszahlen aufbauen, bei denen aber zufällige Schwankungen durch die Bildung von Durchschnittsgrößen aus vergangenen Beobachtungsperioden ausgeschaltet werden.[2)] Ihr Hauptanwendungsgebiet findet die Normalkostenrechnung in der Vorkalkulation zum Zweck der Angebotspreisbildung.[3)] Auch für die Ermittlung von Normalkosten läßt sich die schrittweise Regressionsanalyse verwenden, die aus den Istkosten der Vergangenheit mit Hilfe der Methode der kleinsten Quadrate Durchschnittswerte, die Regreßwerte, bildet. Natürlich müssen genügend viele Beobachtungswerte vorliegen, um zufällige, außerordentliche Abweichungen nach

---

1) Vgl. auch G. MATT: Die schrittweise Regressionsanalyse,., a.a.O., S. 263.

2) P. NOWAK, a.a.O., S. 63; H. VORMBAUM, a.a.O., S. 41; G. WÖHE: Einführung in die Allgemeine Betriebswirtschaftslehre. 8. unveränd. Aufl., Berlin und Frankfurt a.M. 1968. S. 699.

3) H. VORMBAUM, a.a.O., S. 41. So wird z.B. in der Bauindustrie bei der Kalkulation des Angebotspreises in der Regel mit Normalkosten gerechnet.

oben oder unten nicht entscheidend ins Gewicht fallen zu lassen.

Im übrigen gelten für die Ermittlung von Normalkosten durch die schrittweise Regressionsanalyse die für die Aufstellung von Kostenbudgets getroffenen Aussagen sinngemäß.

5.3.4 Multikollinearität bei multiplen Analysen

Bei der Durchführung von multiplen linearen Regressionsanalysen - zu dieser Gruppe gehört auch die schrittweise Regressionsanalyse - d.h. bei Analysen, in denen die gleichzeitige Abhängigkeit einer Erscheinung von mehreren anderen Erscheinungen unter der Annahme einer linearen Verbundenheit untersucht wird, kann das Problem der Multikollinearität von Variablen auftauchen. Von M u l t i k o l l i n e a r i t ä t spricht man, wenn nicht nur zwischen der abhängigen und den unabhängigen Variablen Beziehungen bestehen, sondern wenn auch zwischen den unabhängigen Variablen untereinander eine lineare Korrelation vorliegt.[1)]

Diese Erscheinung kann zu schwierigen Problemen bei der Ermittlung der partiellen Regressions- und Korrelationskoeffizienten führen und es in manchen Fällen sogar unmöglich machen, die Einflüsse der verschiedenen Variablen auf die Zielgröße zu entwirren und eine präzise Schätzung ihrer getrennten Effekte zu erhalten.[2)] Man kann in diesen Fällen wohl den errechneten Wert für die abhängige Variable y' und auch den multiplen Korrelationskoeffizienten als

1) E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 115 f. u. S. 210.; G. TINTNER, a.a.O., S. 259.

2) A.S. GOLDBERGER: Topics in Regression Analysis. New York und London 1968. S. 80.; R.E. JENSEN, a.a.O., S. 271.; G.J. BENSTON, a.a.O., S. 666.

Schätzwerte verwenden,[1] alle partiellen Parameter jedoch, die den von anderen Wirkungen isolierten Einfluß einer Variablen auf die Zielgröße wiedergeben sollen, entsprechen nicht der "wahren" Relation zwischen der abhängigen und der unabhängigen Veränderlichen. Gerade diese partiellen Größen sind es aber, die für den Kostenrechner von besonderer Bedeutung sind, wie auch das im vorigen Abschnitt zitierte Beispiel der Erhöhung des Sintereinsatzes gezeigt hat.[2]

Wenn die Multikollinearität in offener Form auftritt, d.h. "wenn zumindest eine der unabhängigen Veränderlichen mit einer anderen unabhängigen Veränderlichen durch eine lineare Funktion verbunden ist,"[3] läßt sich die Regressionsfunktion nicht eindeutig bestimmen, da die Normalgleichungen auf Grund der funktionalen Beziehung zwischen den unabhängigen Variablen nicht lösbar sind.[4] Bei der Berechnung der Regressions- und Korrelationskoeffizienten mit Hilfe von Matrizen zeigt sich, daß die zur Ermittlung der partiellen Regressionskoeffizienten aufgestellte Matrix singulär ist und deshalb keine inverse Matrix existiert.[5]

Häufiger als die offene Form tritt die versteckte Form der Multikollinearität auf, bei der sich zum Beispiel infolge von Beobachtungs- oder Meßfehlern keine exakte lineare funktionale Beziehung zwischen unabhängigen Veränderlichen

---

1) Vgl. O. ANDERSON, a.a.O., S. 222 f.

2) Vgl. auch den Abschnitt 5.4, S. 118 ff. Hier wird ausführlich dargestellt, weshalb gerade die partiellen Regressionskoeffizienten für die Kostenrechnung von Bedeutung sind.

3) E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 211.

4) G. TINTNER, a.a.O., S. 46.; E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 211.

5) Zur Matrizenrechnung vgl. A. KLINGST, a.a.O., S. 9 ff.

erkennen läßt, die aber gleichwohl vorhanden und lediglich von den Ungenauigkeiten bei der Datenerfassung verschleiert ist.[1] In diesem Fall lassen sich die Regressionskoeffizienten wohl rechnerisch ermitteln, geben jedoch die Beziehungen zwischen den Erscheinungen nicht in korrekter Form wieder.

Um den Einfluß der Multikollinearität zu eliminieren, genügt es nicht, durch Transformationen der unabhängigen Variablen - zum Beispiel durch die Verwendung von x_1 und $(x_2 - x_1)$ als unabhängige Variable anstelle von x_1 und x_2, weil das erste Paar weniger korreliert ist als das zweite - das gesamte Problem zu verlagern. Die auf Grund von x_1 und $(x_2 - x_1)$ errechneten Regressionskoeffizienten sind nämlich nicht mit denen identisch, die sich unter Verwendung von x_1 und x_2 ergeben hätten.[2]

Um den Einfluß der versteckten Form der Multikollinearität zu messen, sind in der Literatur verschiedene Verfahren vorgeschlagen worden, die sich hauptsächlich im Grad ihrer Genauigkeit unterscheiden, so zum Beispiel die von R. FRISCH entwickelte, graphisch durchgeführte Konfluenzanalyse, die Kammlinienanalyse, die von G. TINTNER vorgeschlagene Eigenwertanalyse[3] oder die bei MENGES beschriebene Büschelkartenanalyse.[4] Daneben sind in der Literatur zwei mehr globale Methoden zur Messung der Multikollinearität bekannt, von denen eine sich der Differenz zwischen dem multiplen

---

1) A.S. GOLDBERGER, a.a.O., S. 80.; E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 211.

2) A.S. GOLDBERGER, a.a.O., S. 87.

3) Zur Darstellung und Kritik dieser Verfahren vgl. z.B. E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 213.; G. TINTNER, a.a.O., S. 260.

4) G. MENGES: Ökonometrie. (Die Wirtschaftswissenschaften. Reihe B, Volkswirtschaftslehre, Nr. 20, Hrsg. E. Gutenberg.) Wiesbaden 1961. S. 144 ff.

Bestimmtheitsmaß und der Summe der einfachen Bestimmtheitsmaße als Gradmesser bedient.[1] Nennen wir dieses Maß M_1, so können wir schreiben:

$$M_1 = d_{y.12...m} - \sum_{k=1}^{m} d_{yk}.$$

Da bei fehlender Multikollinearität diese Differenz gleich null sein muß, kann man aus der Höhe einer eventuell auftretenden Differenz auf den Grad der Multikollinearität schließen.[2] Die andere Methode gewinnt das Maß für den Grad der Multikollinearität aus der Aufspaltung des Zählers des multiplen Bestimmtheitsmaßes.[3]

Wegen der lediglich globalen Aussagen dieser Maßgrößen sind bei starker Verbundenheit von unabhängigen Variablen - wenn man zum Beispiel für M_1 einen relativ hohen Wert errechnet hat - stets noch zusätzlich einfache Korrelationskoeffizienten zwischen den Einflußgrößen zu ermitteln. Auf Grund ihrer Größe ist dann zu entscheiden, ob einzelne unabhängige Veränderliche zweckmäßigerweise bei der Bestimmung der multiplen Regressionsgleichung außer Ansatz bleiben sollen.[4]

Die Technik der schrittweisen Regressionsanalyse bietet in diesem Zusammenhang einen wichtigen Vorteil gegenüber der herkömmlichen Vorgehensweise. Sie ermöglicht nämlich durch

---

1) Multikollinearität ist, wie GOLDBERGER ausführt, nämlich eine Frage des G r a d e s und nicht eine Frage des "entweder - oder". Offene Multikollinearität ist deshalb lediglich ein Spezialfall der allgemeinen Multikollinearität. Vgl. A.S. GOLDBERGER, a.a.O., S. 80.

2) Zum Beweis dieses Satzes vgl. E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 212.

3) Einzelheiten vgl. bei E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 212 f. Zum Bestimmtheitsmaß vgl. S. 27 ff.

4) Ebenda. S. 213.

die Eliminierung von statistisch nicht gesicherten Koeffizienten[1] eine Auswahl der für die Vorhersagegleichung "richtigen" unabhängigen Variablen. Wenn eine bisher in der sog. "Zwischenregressionsgleichung" enthaltene Einflußgröße in einem folgenden Schritt ihre statistische Sicherheit verliert - ihr t- oder F-Wert sinkt unter den geforderten kritischen Wert für die Signifikanzprüfung - weil zwischen ihr und der in diesem Schritt in die Regression aufgenommenen Variablen eine Korrelation besteht, dann wird die Ausgangsvariable aus der Gleichung wieder eliminiert.[2] Auf diese Weise kann man zu einer endgültigen Vorhersagegleichung gelangen, in der nur statistisch gesicherte, d.h. in diesem Fall weitgehend voneinander unabhängige Variable enthalten sind.

Bei der Untersuchung der Abhängigkeiten zwischen Kosten oder ihren Mengenkomponenten und den Einflußfaktoren liegt häufig der Verdacht auf Existenz von Multikollinearität nahe, da die kostenbeeinflussenden Faktoren oft miteinander in Beziehung stehen. Nicht immer bietet sich aber wie in dem zitierten Beispiel der Bestimmung der Komponenten des Koksverbrauchs[3] die Möglichkeit, aus einer großen Anzahl von unabhängigen Variablen die wichtigen auswählen zu können und alle mit ihnen verbundenen Größen bei der Vorhersagegleichung außer Ansatz zu lassen. In manchen Fällen m ü s s e n alle unabhängigen Variablen in den Gleichungsansatz aufgenommen werden, wenn nämlich das Ziel der Regressionsanalyse in der Aufteilung einer bestimmten Kostensumme auf einzelne Produkte besteht. Man denke hier bei-

---

1) Vgl. auch S. 105 f.

2) Ebenda.

3) Vgl. S. 107 ff.

spielsweise an die Ermittlung von Grenzkosten im Fall der verbundenen Produktion mit loser Kopplung, die über die Bestimmung der partiellen Regressionskoeffizienten erfolgen kann.[1] Die Grenzkosten können nur dann durch die multiple Regressionsanalyse korrekt ermittelt werden, wenn die mengenmäßige Produktion der in dieser Analyse betrachteten Güter unabhängig voneinander ist, d.h. wenn durch die Herstellung einer bestimmten Anzahl des Produktes 1 nicht zugleich eine bestimmte Produktionsmenge der Erzeugnisse 2, 3, ..., m festgelegt ist. Nur in diesem Fall kann man die Existenz von Multikollinearität von vornherein ausschließen.[2]

BENSTON weist darauf hin, daß die Gefahr für das Auftreten von Multikollinearität bei Kostenuntersuchungen um so größer ist, je weiter die betrachteten Kostenarten zu aggregierten Kosten zusammengefaßt werden. Durch Aufspaltung der Gesamtkosten in mehrere "Untergruppen", die voneinander unabhängig sein müssen, kann jedoch eine Verbesserung des Ausgangsdatenmaterials erreicht werden.[3] Eine wichtige Aufgabe fällt hierbei der Verteilung - und damit Disaggregation - der Kosten auf Kostenstellen zu, bevor daran anschließend eine multiple Regressionsanalyse für die Kosten jeder einzelnen Stelle durchgeführt wird. Als Nebeneffekt erreicht man zugleich eine Reduzierung der Anzahl der unabhängigen Variablen.[4]

---

1) Zur ausführlichen Darstellung dieses Problems vgl. den Abschnitt 5.4, S. 118 ff.

2) Diese Aussage gilt entsprechend für das im Abschnitt 4.3 zitierte Beispiel der Verrechnung des Stromverbrauchs auf die Kostenträger.

3) Vgl. G.J. BENSTON, a.a.O., S. 666.

4) Ebenda.

Obwohl auch diese Aufteilung der Kosten auf Kostenstellen viele Probleme und Ungenauigkeiten mit sich bringt,[1] erscheint eine solche Voruntersuchung und Verteilung der Gesamtkosten v o r der eigentlichen Analyse als wirkungsvollste Methode, den Grad der Multikollinearität niedrig zu halten, so daß man zu einigermaßen exakten Schätzwerten für die Regressionskoeffizienten gelangt. Eine Reduzierung der Standardfehler der Koeffizienten - und damit eine Verringerung ihres Schwankungsbereichs - durch eine Erhöhung des Stichprobenumfangs wird bei Kostenuntersuchungen nur in seltenen Fällen möglich sein, da in der Regel nicht genügend Beobachtungsperioden oder Untersuchungsobjekte zur Verfügung stehen oder aufgezeichnet wurden[2] und eine beliebige Erhöhung des Stichprobenumfangs durch Experimente - wie es in den Naturwissenschaften häufig möglich ist - für kostenrechnerische Untersuchungen ausscheidet.[3]

Zusammenfassend kann man feststellen, daß das Problem der Multikollinearität die sinnvolle Anwendung multipler Regressionen für Kostenuntersuchungen erheblich erschweren kann, insbesondere dann, wenn auf Grund der Aufgabenstellung a l l e unabhängigen Variablen in der Regression enthalten sein m ü s s e n und man nicht durch eine schrittweise Regressionsanalyse mit Tests der Koeffizienten einzelne Variable aus der Vorhersagegleichung eliminieren kann. In diesen Fällen sollten, um unsinnige Ergebnisse zu vermeiden, Tests auf Multikollinearität[4] durchgeführt werden und bei positivem Ergebnis sollte versucht

---

1) Vgl. das Problem der Wahl der optimalen Schlüsselgröße für die Verrechnung von Kostenstellengemeinkosten. Vgl. Abschnitt 4.4, S. 58 ff.

2) Vgl. auch R.E. JENSEN, a.a.O., S. 271 und S. 229 ff. dieser Arbeit.

3) Hinzu tritt der Gesichtspunkt, daß auch eine Stabilität des Ausgangsdatenmaterials über eine längere Zeitspanne in der Regel nicht gegeben ist.

4) Vgl. S. 113 f.

werden, durch Umformung, d.h. in der Regel durch Aufspaltung des Datenmaterials, den Grad der Multikollinearität zu senken.

5.4 Kostenauflösung zur Gewinnung von Unterlagen für kurzfristige Entscheidungen im Produktionsbereich

Schon im Rahmen der Ermittlung von Budgetkosten wurde auf die Bedeutung der Kostenrechnung als Instrument zur Bereitstellung von Unterlagen für kurzfristige Entscheidungen der Unternehmensleitung hingewiesen.[1] Die hierfür erforderlichen Daten, die auch für besondere Kostenrechnungsverfahren (zum Beispiel das "Direct Costing"[2] oder die Grenzplankostenrechnung[3]) benötigt werden, können von der üblicherweise durchgeführten Kostenarten-, Kostenstellen- und Kostenträgerrechnung nicht ermittelt werden, sondern müssen außerhalb der eigentlichen Betriebsabrechnung durch Sonderrechnungen gewonnen werden. Es handelt sich hierbei im wesentlichen um die Ermittlung von Grenzkosten eines Produktes, einer Produktart oder eines Auftrags. Die Bestimmung der Grenzkosten setzt "eine Aufspaltung der Kostensumme in fixe, von der Beschäftigung mindestens in Intervallen unabhängige, und variable, mit der Beschäftigung schwankende Kosten"[4] voraus, wobei die variablen Kosten die Grenzkosten einer zusätzlichen Einheit bestimmen. Die-

1) Vgl. S. 97 ff.

2) Vgl. zum Begriff des "Direct Costing" z.B. H.-H. BÖHM u. F. WILLE: Deckungsbeitragsrechnung und Programmoptimierung. 2. völlig überarb. u. erw. Aufl. des Werkes "Direct Costing und Programmplanung". München 1965. S. 17 ff.

3) Vgl. z.B. P. RIEBEL: Das Rechnen mit Einzelkosten und Deckungsbeiträgen. "Zeitschrift für handelswissenschaftliche Forschung". Neue Folge. 11 (1959) 3, S. 213 ff.

4) D. SCHNEIDER, Kostentheorie ..., a.a.O., S. 679.

se Aufspaltung der Kosten ist unter dem Begriff K o s t e n a u f l ö s u n g bekannt.

In amerikanischen Veröffentlichungen findet man zahlreiche Beschreibungen dieses Verfahrens, wobei sich insbesondere in neueren Zeitschriftenartikeln Anwendungen der Regressions- und Korrelationsanalyse feststellen lassen, ein Trend, der sicherlich durch die Möglichkeit zur elektronischen Datenverarbeitung gefördert worden ist.[1)]

In den folgenden Abschnitten soll zunächst ein allgemeiner Überblick über das Verfahren der statistischen Kostenauflösung gegeben werden. Daran anschließend wird am Beispiel der Sortimentsplanung bei verbundener Produktion und am Beispiel einer Entscheidung für Selbstherstellung oder Fremdbezug die Grenzkostenermittlung durch einfache und durch partielle Regressionskoeffizienten ausführlicher dargestellt.

5.4.1 Grundlagen der statistischen Kostenauflösung

Zur praktischen Durchführung der Kostenauflösung werden in der Literatur verschiedene Verfahren beschrieben,[2)] von denen im Rahmen dieser Arbeit die sog. "statistische Kostenauflösung" besondere Beachtung verdient. Auf die anderen vorgeschlagenen Methoden, so zum Beispiel die von SCHMALENBACH propagierte "mathematische Kostenauflösung"[3)] soll hier dagegen nicht weiter eingegangen werden.

---

1) Vgl. K. WEBER: Besonderheiten der amerikanischen Kostenrechnung. "Zeitschrift für Betriebswirtschaft". 38 (1968) 11, S. 837 ff., hier S. 851.

2) Vgl. z.B. D. SCHNEIDER, Kostentheorie ..., a.a.O., S. 680 ff.

3) Zur Kritik dieses Verfahrens vgl. z.B. D. SCHNEIDER, Kostentheorie ..., a.a.O., S. 680.; J. MALETZ: Kostenauflösung. "Zeitschrift für handelswissenschaftliche Forschung". 20 (1926) 7, S. 293 ff., hier S. 302 ff.; K. MELLEROWICZ: Kosten und Kostenrechnung. Bd. I: Theorie der Kosten. 4. durchges. Aufl. Berlin 1963. S. 357 ff.

Die statistische Kostenauflösung geschieht in der Regel mit Hilfe der nach der Methode der kleinsten Quadrate durchgeführten Regressionsanalyse,[1] wobei die in der Vergangenheit beobachteten Kostensummen einer Kostenart in einer Kostenstelle die abhängige und die Beschäftigung dieser Kostenstelle - gemessen in Maschinenlaufstunden, Arbeitsstunden, Anzahl der gefertigten Produkte oder ähnlichen Größen - die unabhängige Variable darstellen. Da die Kosten bei dem üblichen Verfahren lediglich in Abhängigkeit von einer Größe untersucht werden, kommt nur eine e i n f a c h e Regressionsanalyse in Betracht. Sie wird ausschließlich als lineare Form mit der Ausgangsgleichung $y' = a + bx$ durchgeführt, "wobei a das fixe Element und b den Grad der Proportionalität darstellt".[2] Bei Annahme einer geradlinig verlaufenden Kostenkurve entspricht b den Grenzkosten,[3] da b definiert ist als der Betrag, um den die abhängige Variable (die Kosten) sich ändert bei Variation der unabhängigen Variablen (Beschäftigung) um eine Einheit.

Man erhält auf Grund des beobachteten Datenmaterials wieder die schon mehrfach erwähnte Punktwolke im zweidimensionalen Koordinatensystem, durch die nach der Methode der kleinsten Quadrate die lineare Regressionsfunktion mit den Parametern a und b gelegt wird.

---

1) Daneben nennt MATZ noch die "Hoch-Tiefpunkt-Methode" und "Streubildanalysen". Vgl. A. MATZ, a.a.O., S. 87.

2) A. MATZ, a.a.O., S. 91.

3) Vgl. auch C. NEBELUNG: Statistische Ermittlung der fixen und variablen Kosten in Betrieben der Großserienfertigung. "Zeitschrift für Betriebswirtschaft". 20 (1950) 7, S. 417 ff. A. MUIR: Marginal Cost. "Applied Statistics. A Journal of the Royal Statistical Society". 7 (1958) 2, S. 96 ff., hier S. 99.

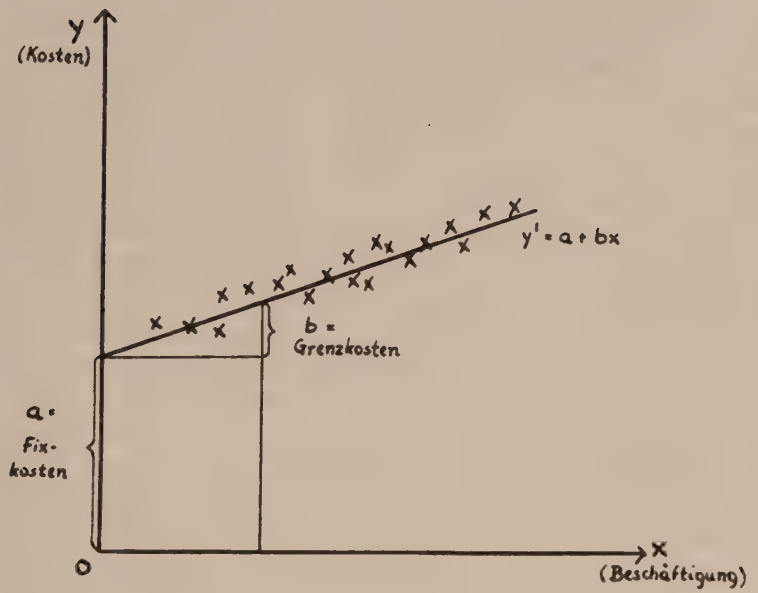


Abb. 5.3: Statistische Kostenauflösung über die Bestimmung der Regressionsparameter.

Auf ein Zahlenbeispiel zur Berechnung der Regressionsparameter soll hier verzichtet werden, da das Verfahren aus früheren Darstellungen bekannt ist und sich auch in der Literatur Beispiele für die praktische Ermittlung der Grenzkosten finden.[1)]

1) Vgl. z.B. A. MATZ, a.a.O., S. 92 f.

Vielmehr soll an dieser Stelle auf einige wichtige Gesichtspunkte für die Beurteilung der statistischen Kostenauflösung mit Hilfe der einfachen Regressionsanalyse hingewiesen werden. Da die durch die Regressionsrechnung ermittelten Grenzkosten auf Vergangenheitszahlen basieren, wird man sie zur Lösung praktischer Planungsfragen der Gegenwart oder Zukunft in der Regel nicht ohne vorherige Bereinigungen übernehmen dürfen.[1] Das kann dazu führen, daß die Ceteris-paribus-Klausel, deren Einhaltung zur Ermittlung des Zusammenhangs von Kosten und Beschäftigungsgrad notwendig war - Ausschaltung aller Einflüsse außer dem der betrachteten Kosteneinflußgröße Beschäftigung - wieder aufgehoben werden muß, wenn zum Beispiel Preisentscheidungen getroffen werden sollen. Denn um die Frage beantworten zu können, ob der für einen Auftrag gebotene Preis über den Grenzkosten dieses Auftrags liegt, sind im Kalkulationszeitpunkt möglicherweise aufgetretene Änderungen der Kostengüterpreise zu berücksichtigen.[2] Diese Aussage gilt entsprechend für alle anderen Kosteneinflußgrößen.

Das Dilemma besteht darin, einerseits durch die Aufnahme möglichst vieler Vergangenheitswerte einen relativ großen Stichprobenumfang zu erhalten, um zu statistisch zuverlässigeren Schätzwerten zu gelangen,[3] andererseits aber nicht zu weit in die Vergangenheit zurückzugehen, da die ermittelten Grenzkosten ja gegenwärtige Situationen wiedergeben sollen.[4] In manchen Fällen bietet sich die Möglichkeit an, das ältere Datenmaterial durch Berichtigungen

---

1) Vgl. auch S. 129, wo auf diese verzerrenden Einflüsse ausführlicher eingegangen wird.

2) D. SCHNEIDER, Kostentheorie ..., a.a.O., S. 682.

3) Vgl. A. MUIR, a.a.O., S. 99.; Vgl. auch Abschnitt 5.6, S. 141 ff.

4) Vgl. A. MUIR, a.a.O., S. 99.

und Umformungen zu "erneuern", d.h. auf Gegenwartswerte umzustellen.[1] Es zeigt sich hier wieder die schon früher - so im Rahmen der Ermittlung der Sollkostenkurve in der Plankostenrechnung[2] - erwähnte Problematik der Gewinnung von Planungszahlen aus beobachteten Vergangenheitswerten.

Darüber hinaus entsteht auch wieder das Problem der Extrapolation der Regreßwerte, d.h. im Fall der Kostenauflösung die Verlängerung der Regressionslinie in Richtung auf den Ursprung des Koordinatensystems, um die Regressionskonstante a bei einem Wert von $x = 0$ als fixe Kosten interpretieren zu können.[3] Es wurde bereits ausführlich dargestellt, aus welchen Gründen die auf Grund eines bestimmten, eng begrenzten Schwankungsbereichs für die unabhängige Variable ermittelte Regressionskonstante nicht ohne weiteres als Fixkosten gedeutet werden darf.[4] Nach Ansicht des Verfassers kann diese Schwierigkeit dadurch ausgeklammert werden, daß im Rahmen einer m u l t i p l e n Regressionsanalyse auf die Interpretation der Regressionskonstanten und damit auf die rechnerische Bestimmung der Fixkosten ganz verzichtet wird, da sie für die zu treffenden kurzfristigen Entscheidungen nicht relevant sind. Man sollte sich also darauf beschränken, die partiellen Regressionskoeffizienten als Grenzkosten der verschiedenen Leistungseinheiten zu deuten, wie dies im folgenden Abschnitt bei der Grenzkostenbestimmung für Zwecke der Sortimentsplanung bei verbundener Produktion mit loser Kopplung geschieht.

---

1) A. MUIR, a.a.O., S. 99.

2) Vgl. S. 86 ff.

3) Vgl. S. 50 ff.

4) Vgl. insbesondere die Abb. 4.1 und 4.2 auf S. 53 und S. 54.

5.4.2 Sortimentsplanung bei verbundener Produktion

5.4.2.1 Problemstellung

Da bei der weitaus größten Zahl aller Betriebe die Produktion nicht auf die Herstellung eines Gutes beschränkt ist, sondern in der Regel eine Vielzahl von Produkten erzeugt wird, bildet die Planung des Produktionssortiments einen wichtigen Bestandteil der unternehmerischen Gesamtplanung. Unter Produktionssortiment verstehen wir dabei die Kombination der Arten und Mengen von Absatzgütern.

Wie die Theorie lehrt, ist bei Betrieben, die nach dem Gesetz des gewinnmaximalen Prinzips handeln, das Gewinnmaximum dann erreicht, wenn die Grenzkosten eines Produktes gleich dem Grenzerlös oder dem Preis dieses Produktes sind.[1] Um dieses theoretische Kriterium mit Zahlenwerten ausfüllen zu können, müssen demnach zunächst einmal die Grenzkosten eines Produktes bekannt sein, d.h. die zusätzlichen Kosten für die Erstellung einer weiteren Leistungseinheit.[2] Abgesehen von Schwierigkeiten, die sich bei der praktischen Ermittlung durch Ungenauigkeiten bei der Aufzeichnung, durch Bildung von Durchschnittswerten und infolge von Schlüsselungen ergeben, werden diese Grenzkosten noch verhältnismäßig einfach zu ermitteln sein, wenn ein Betrieb mehrere Erzeugnisse herstellt, deren Fabrikation nicht miteinander verbunden ist.[3] Diese Betriebe, die mit sog. "isolierter Fertigung" arbeiten, bilden jedoch die Ausnahme, so daß in aller Regel eine "verbundene Produktion" vorliegt.[4] Man versteht darunter eine Form der Produktion, bei der "meh-

1) Vgl. z.B. E. GUTENBERG, a.a.O., (1970), S. 455.; G. WÖHE, a.a.O., S. 280 ff.

2) E. SCHMALENBACH, a.a.O., S. 77.

3) E. GUTENBERG: Grundlagen der Betriebswirtschaftslehre. 1. Band: Die Produktion. 8./9. Aufl. Berlin/Göttingen und Heidelberg 1963. S. 116.

4) Ebenda. S. 117.

rere Arten von Erzeugnissen unter weitgehender Benutzung ein und desselben Produktionsapparates hergestellt werden."[1] In diesem Fall sind die Grenzkosten einer Produktart nicht nur von der hergestellten Menge dieses Produktes, sondern auch von den Produktionshöhen der übrigen Erzeugnisse abhängig. "Sie ergeben sich jeweils für die einzelnen Produktarten, wenn man die Ausbringung einer Produktart um eine Einheit erhöht und die übrigen Ausbringungen konstant hält."[2]

Diese Aussagen können sich selbstverständlich nur auf Fälle verbundener Produktion mit l o s e r Kopplung beziehen, d.h. auf die Herstellung von Gütern, bei denen man Einfluß auf die Zusammensetzung (das Mengenverhältnis) des Produktionssortiments nehmen kann.[3] Fällt dagegen eine bestimmte Güterkombination auf Grund irgendwelcher natürlicher Gesetzmäßigkeiten notwendigerweise in einer ganz bestimmten, nicht variierbaren Kombination an (Kuppelprodukte), dann lassen sich lediglich die Gesamtkosten dieser Kombination, nicht aber die Grenzkosten eines Teilproduktes ermitteln.[4] Aus diesem Grund werden die Fälle von Kuppelproduktion aus den folgenden Erörterungen ausgeschlossen.

---

1) E. GUTENBERG, a.a.O., (1963), S. 117.

2) Ebenda.

3) Vgl. auch J.S. CHIU und D.T. DE COSTER: Multiple Product Costing by Multiple Correlation Analysis. "The Accounting Review". 41 (1966) 4, S. 673 ff., hier S. 678.; A.N. LORIG: Joint Cost Analysis as an Aid to Management. "The Accounting Review". 30 (1955) 4, S. 634 ff., hier S. 635.

4) Teilweise auch als "starre" bzw. "elastische" Kopplung bezeichnet, so K. KÄFER, a.a.O., S. 357.

5.4.2.2 Ermittlung der Grenzkosten durch partielle Regressionskoeffizienten

Wie können nun die Grenzkosten von Produkten, die in loser, d.h. variierbarer Mengenzusammensetzung erzeugt werden, in der Praxis ermittelt werden? Im folgenden soll ein Versuch unternommen werden, die Grenzkosten mit Hilfe der multiplen Regressionsanalyse zu berechnen, indem die Grenzkosten jeder Produktart durch die einzelnen partiellen Regressionskoeffizienten dargestellt werden.

Es soll bereits an dieser Stelle darauf hingewiesen werden, daß auch die Regressionsanalyse nicht d i e exakten Grenzkosten ermitteln kann, da stets Ungenauigkeiten, Meß- und Kontierungsfehler, willkürliche Schlüsselungen und ähnliche verzerrende Einflüsse auf die Ausgangsdaten für die Analyse einwirken können.[1] Der Verfasser ist aber der Meinung, daß die im folgenden gezeigte Lösung zu einer vernünftigen und relativ einfachen Ermittlung von Grenzkosten für die Sortimentsplanung beitragen kann.

Zur Verdeutlichung der Vorgehensweise soll wieder ein Rechenbeispiel dienen.[2] Der Einfachheit halber wollen wir davon ausgehen, daß die Produktarten vorgegeben sind und lediglich Variationen im Mengenverhältnis dieser Produkte vorgenommen werden. Die Gesamtkosten für die Produktion verschiedener Mengen der drei betrachteten Produktarten werden mit y, die Mengen der Produktarten selbst mit x_1, x_2 und x_3 bezeichnet. Wir nehmen an, daß wir während zehn Beobachtungsperioden folgende Werte für die abhängige und die unabhängigen Variablen aufgezeichnet haben:

---

1) Auf die Problematik von Interkorrelationen zwischen den unabhängigen Variablen wurde bereits eingegangen. Die besonderen Schwierigkeiten bei der Analyse von Zeitreihen werden im Abschnitt 7.1, S. 215 ff. behandelt.

2) Die Zahlen wurden unverändert übernommen von J.S. CHIU u. D.T. DE COSTER, a.a.O., S. 676.

Periode	y_i a)	x_{1i}	x_{2i}	x_{3i}
1	80,8	232	87	390
2	77,8	221	172	262
3	76,4	258	125	313
4	67,5	300	52	300
5	83,5	198	151	349
6	80,0	183	92	489
7	37,9	181	33	131
8	88,7	159	111	441
9	84,5	197	209	228
10	105,3	145	234	410
$\sum$	782,4	2074	1266	3313

a) Kostenangabe in 1000 Geldeinheiten

Tabelle 5.1: Gesamtkosten y_i und Produktionsmengen x_{1i}, x_{2i}, x_{3i} von drei Erzeugnissen für 10 Beobachtungsperioden.

Legt man eine Regressionsgleichung der Form

$$y' = a_{y.123} + b_{y1.23}x_1 + b_{y2.13}x_2 + b_{y3.12}x_3$$

zugrunde und ermittelt die Werte nach der Methode der kleinsten Quadrate, so erhält man für die Regressionskonstante und die partiellen Regressionskoeffizienten folgende Werte:[1)]

$$a_{y.123} = 10193,15,$$

$$b_{y1.23} = 48,67,$$

$$b_{y2.13} = 193,94 \text{ und}$$

$$b_{y3.12} = 100,81.$$

1) Entnommen aus J.S. CHIU u. D.T. COSTER, a.a.O., S. 676.

Gemäß der Definition für partielle Regressionskoeffizienten kann man sagen, daß die Kosten sich um 48,67 Geldeinheiten erhöhen, wenn eine Einheit des Gutes 1 mehr produziert wird, unter der Voraussetzung, daß die Produktion der Güter 2 und 3 unverändert bleibt.[1] Diese Größe ist aber nichts anderes als die Grenzkosten des Produktes 1, die ja auch definiert sind als Kostenzunahme bei Erhöhung der Produktion des Gutes 1 um eine Einheit bei Konstanz der Produktionsmengen aller übrigen Erzeugnisse. Auf diese Weise lassen sich für einen gegebenen Datensatz die Grenzkosten verschiedener Produkte relativ leicht mit Hilfe der multiplen Regressionsanalyse ermitteln.

Es darf bei diesem Vorgehen nicht übersehen werden, daß die partiellen Regressionskoeffizienten auf Grund einer Stichprobe aus einer übergeordneten Grundgesamtheit ermittelt wurden und deshalb nicht unbedingt mit den "wahren" Regressionskoeffizienten der Grundgesamtheit, d.h. den tatsächlichen Grenzkosten, identisch sein müssen. Die statistischen Methoden geben uns aber die Möglichkeit, die Zuverlässigkeit unserer Schätzungen zu überprüfen, indem wir Konfidenzintervalle für die Regressionsparameter bilden, in denen mit einer bestimmten Wahrscheinlichkeit die wahren Werte der Grundgesamtheit zu suchen sind.[2]

---

1) Für x_2 und x_3 gilt diese Aussage entsprechend. Man erhält in diesem Fall die Grenzkosten des Produktes 2 bzw. 3.

2) Zur ausführlichen Darstellung der Bildung von Konfidenzintervallen vgl. den Abschnitt 5.6, S. 141 ff.

5.4.2.3 Beurteilung der regressionsanalytischen Grenzkostenermittlung

Der wesentliche Vorteil der Ermittlung von Grenzkosten bei verbundener Produktion durch die multiple Regressionsanalyse liegt in der Tatsache, daß durch die partiellen Regressionskoeffizienten die Grenzkosten a l l e r betrachteten Produkte s i m u l t a n bestimmt werden und dadurch die Kostenstruktur klar erkennbar wird.[1] Deshalb entspricht nach Meinung des Verfassers dieses Verfahren auch der von der betriebswirtschaftlichen Theorie geforderten Beachtung der Ceteris-paribus-Bedingung, da die Grenzkosten eines Produktes sich aus der Veränderung der Gesamtkosten ergeben, wenn die Ausbringung dieses Produktes um eine Einheit erhöht und die Produktionszahlen der übrigen Erzeugnisse konstant gehalten werden.[2] Gerade dies geschieht durch die Ermittlung der p a r t i e l l e n Regressionskoeffizienten.[3]

Auch bei der Anwendung der Regressionsanalyse auf Probleme der Sortimentsplanung sind die schon früher erwähnten Vorbehalte zu machen. Es muß bei der Analyse, die sich auf Datenmaterial aus mehreren v e r g a n g e n e n Beobachtungsperioden stützt, und bei der anschließenden Interpretation der Ergebnisse vorausgesetzt werden können, daß sich zwischenzeitlich keine Änderungen im Produktionsablauf, sei es durch Änderung der Produktionsmethoden, durch den Einsatz anderer Maschinen, anderer Arbeitskräfte oder an-

1) Vgl. auch J.S. CHIU u. D.T. DE COSTER, a.a.O., S. 677.
2) Vgl. E. GUTENBERG, a.a.O., (1963), S. 117.
3) Vgl. auch S. 24 f.

deren Materials usw. vollzogen haben.[1] Andernfalls können die gewonnenen Ergebnisse nur dann weitere Verwendung finden, wenn es gelingt, durch Korrekturen diesen Veränderungen Rechnung zu tragen.

Eine Extrapolation der statistischen Relation ist - wie bei allen anderen Regressionsmodellen - nicht ohne weiteres zulässig;[2] lediglich innerhalb des Schwankungsbereichs der Beobachtungswerte dürfen Interpolationen vorgenommen werden, da diese "Zwischenwerte" ihren Niederschlag in der ermittelten Regressionsgleichung über die Beobachtungswerte gefunden haben.

Auch bei Berücksichtigung der Problematik, die alle Berechnungsverfahren beinhalten, mit denen auf Grund von in der betrieblichen Praxis aufgezeichneten Daten betriebswirtschaftliche Analysen durchgeführt werden, kann die multiple Regressionsanalyse als wertvolles Hilfsmittel bei der Grenzkostenermittlung von verbundenen Produkten zum Zweck der Sortimentsplanung angesehen werden. Auch an dieser Stelle sei der Hinweis gestattet, daß selbst bei umfangreichem Zahlenmaterial die Durchführung der Analyse auf elektronischen Rechenanlagen in sehr kurzer Zeit bewältigt werden kann. Nach einer gewissen Einführungszeit wird diese Analyse zu einer Routineangelegenheit, so daß der Kostenrech-

---

1) Allerdings werden sich in den seltensten Fällen Änderungen im Produktionsablauf völlig vermeiden lassen, da zum Beispiel bei Arbeitnehmern infolge von Urlaub, Krankheit oder ähnlichen Einflüssen Verschiebungen an der Tagesordnung sind. Diese und ähnliche Veränderungen können jedoch bei der Interpretation der Regressionsergebnisse und bei ihrer Anwendung in der Zukunft ohne Berücksichtigung bleiben, da ihre Bedeutung in aller Regel zu gering ist, um das aus einer Vielzahl von Einflüssen resultierende Gesamtergebnis spürbar zu verändern.

2) Vgl. S. 50 ff., S. 89 ff., M. EZEKIEL u. K.A. FOX, a.a.O., S. 138.

ner selbst sich völlig der Aufbereitung und Interpretation des Datenmaterials und der Ergebnisse widmen kann, ohne wertvolle Zeit mit der Durchführung der Rechenoperationen zu verlieren.

5.4.3 "Make-or-buy"-Entscheidungen

Häufig sieht sich ein Unternehmer vor die Frage gestellt, ob es für ihn vorteilhaft ist, bestimmte Güter selbst zu produzieren bzw. bestimmte Leistungen selbst zu erbringen oder von fremden Unternehmen einzukaufen bzw. ausführen zu lassen. Man denke zum Beispiel an elektrischen Strom, an Zubehörteile in der Automobilindustrie, an die Frage in der Bauindustrie, ob Beton auf der Baustelle selbst hergestellt werden soll oder ob man ihn als Fertigbeton von einem anderen Werk bezieht, und ähnliche Fälle mehr. Diese Problematik wird in der angloamerikanischen Literatur mit dem Begriff "make or buy" bezeichnet, der sich auch weitgehend im deutschen Schrifttum und Sprachgebrauch eingebürgert hat.

Um zu zeigen, wie auch in diesem Fall die Regressionsanalyse dem Kostenrechner wertvolle Ergebnisse liefern kann, soll kurz auf ein Beispiel eingegangen werden, das man bei A. MUIR findet und das eindrucksvoll zeigt, welche Vorteile das statistische Verfahren gegenüber einer herkömmlichen Berechnungsmethode für die Grenzkosten bietet.[1)]

1) Vgl. A. MUIR, a.a.O., S. 99 ff. MUIR geht von einer Stichprobe vom Umfang n = 24 aus und unterteilt die abhängige Variable in drei Komponenten: Löhne, andere Betriebskosten und Kohleverbrauch. Auf die Wiedergabe der Zahlenwerte und der Streupunktdiagramme kann hier verzichtet werden, da es dem Verfasser ausschließlich auf die Interpretation der Ergebnisse und den Vergleich mit anderen Methoden der Grenzkostenermittlung ankommt.

Man stelle sich einen Betrieb vor, der seinen Dampfbedarf entweder aus der Erzeugung seines eigenen Kesselhauses oder aus der Lieferung einer naheliegenden fremden Kraftstation decken kann. Es erhebt sich die Frage, ob ein in Zukunft zu erwartender Mehrbedarf an Dampf selbst produziert oder bezogen werden soll. Da das Unternehmen bereits einen fixen Betrag (Grundgebühr) an die fremde Kraftstation zahlt und auch für das eigene Kesselhaus fixe Kosten in jedem Fall entstehen, ist es bei der Entscheidung für eine Alternative lediglich notwendig zu wissen, ob die G r e n z k o s t e n des eigenen Dampfes oder die l a u f e n d e n Kosten des fremden Dampfes höher sind. Dies ist eine typisch k u r z f r i s t i g e Fragestellung, indem man nämlich von einem gegebenen Produktionsapparat ausgeht und lediglich die variablen oder laufenden Kosten miteinander vergleicht.

Indem MUIR die abhängige Variable "Kosten der Dampferzeugung" in drei Komponenten aufteilt, erhält er drei einfache Regressionen zwischen diesen Veränderlichen und der Produktionshöhe des Dampfes, die jeweils die unabhängige Variable bildet. Aufgrund seines Zahlenmaterials gelangt er zu den folgenden Regressionskoeffizienten, die nach ihrer Definition gleich den Grenzkosten des Dampfes sind, und folgenden Standardfehlern der Regressionskoeffizienten:[1)]

---

1) Die Zahlen wurden unverändert von A. MUIR, a.a.O., S. 101 übernommen. Auf die Ermittlung der Standardfehler der Regressionskoeffizienten wird ausführlich im Abschnitt 5.6 eingegangen, vgl. S. 150 ff. Da die Regressionskoeffizienten als Grenzkosten interpretiert werden, sind die Standardfehler der Regressionskoeffizienten ex definitione gleich den Standardfehlern der Grenzkosten.

Kosten-art	Grenzkosten	Standardfehler der Grenzkosten
Löhne	0,01955	0,0021
Betr.Ko.	0,01313	0,0007
Kohle	0,28939	0,0057
$\sum$	0,32207	0,0062

Tabelle 5.2: Grenzkosten und ihre Standardfehler für drei Kostenarten zur Ermittlung der optimalen Alternative zwischen Selbstherstellung und Fremdbezug.

Der Regressionskoeffizient für die Abhängigkeit zwischen Kohleverbrauch und erzeugtem Dampf wurde mit dem Faktor 4,4 multipliziert, dem durchschnittlichen Kohlepreis während der Zeit der Analyse, um aus einer Mengengröße die Wertgröße "Grenzkosten" zu erhalten. Durch die Addition der drei Regressionskoeffizienten erhält man die totalen Grenzkosten für das Produkt "Dampf". Die Addition der Standardfehler der Grenzkosten ergibt nicht den als Summe ausgewiesenen Wert, weil, um zu korrekten Ergebnissen zu gelangen, die Varianzen, d.h. die Quadrate der Standardfehler, addiert werden müssen und aus dieser Summe die Quadratwurzel zu ziehen ist, um die gesamte Standardabweichung zu erhalten.[1)]

Die gesamten Grenzkosten in Höhe von 0,322 Geldeinheiten[2)] werden nun dem für eine Einheit Dampf bei Fremdbezug zu zahlenden Betrag gegenübergestellt, der im Durchschnitt zur Zeit der Analyse 0,342 Geldeinheiten betrug. Um zu testen,

1) Vgl. auch A. MUIR, a.a.O., S. 101.

2) Zur Vereinfachung der folgenden Rechnung wurde der Wert auf drei Stellen hinter dem Komma gerundet.

ob zwischen diesen beiden Größen ein signifikanter Unterschied besteht, d.h. ob sie wesentlich voneinander abweichen, ermitteln wir eine Testgröße z, deren Wert sich aus der Differenz zwischen den Grenzkosten und dem Preis des Dampfes, geteilt durch die Standardabweichung der Grenzkosten errechnet.[1] Setzen wir die ermittelten Werte ein, so erhalten wir:[2]

$$z = \frac{0,342 - 0,322}{0,0062} = \frac{0,020}{0,0062} = 3,23.$$

Aus der Tabelle der Studentverteilung[3] können wir entnehmen, daß dieser z-Wert bei 22 Freiheitsgraden[4] mit einer statistischen Wahrscheinlichkeit von über 99 % auf einen signifikanten Unterschied zwischen den beiden untersuchten Werten - Grenzkosten bei eigener Dampferzeugung und Preis des Dampfes bei Fremdbezug - schließen läßt. Das bedeutet, daß der eigene Dampf billiger produziert als der fremde Dampf gekauft werden kann (Grenzbetrachtung). Wenn also in Zukunft ein Mehrbedarf an Dampf erwartet wird, sollte dieser auf Grund der vorgenommenen Untersuchung im eigenen Kesselhaus erzeugt werden.

Hätte man jedoch die Grenzkosten der Stromerzeugung nicht nach einzelnen Kostenarten getrennt ermittelt, wie dies durch die drei Regressionen geschah - mit anschließender Addition der "Teil-Grenzkosten" zu den gesamten Grenzkosten - sondern nach der üblicherweise angewandten Beziehung: Durch-

---

1) Auf die statistischen Grundlagen dieser Berechnungsmethode wird ausführlich im Abschnitt 5.6 eingegangen, vgl. S. 141 ff.

2) Die Werte wurden unverändert übernommen von A. MUIR, a.a.O., S. 101.

3) Vgl. die Tabelle 4* im Anhang II, S. 287.

4) Die Zahl der Freiheitsgrade errechnet sich nach der Formel $\nu = n - k - 1$, wobei n den Stichprobenumfang und k die Zahl der unabhängigen Variablen in der Regression repräsentiert. Bei $n = 24$ und $k = 1$ erhält man für ν: $\nu = 24 - 1 - 1 = 22$.

schnitt der drei addierten Kostenarten Löhne, Betriebskosten und Kohlekosten geteilt durch die durchschnittliche Dampferzeugung, so hätte man folgenden Wert erhalten:[1)]

$$\frac{M_L + M_B + M_K}{M_D} = \frac{803 + 198 + 663 \cdot 4{,}4}{8423} = 0{,}465.$$

In dieser Gleichung bedeuten:

M_L = Mittelwert der Löhne,

M_B = Mittelwert der Betriebskosten,

M_K = Mittelwert der Kosten für Kohlen[2)] und

M_D = Mittelwert der Dampferzeugung.

Vergleicht man diesen Wert von 0,465 mit den beiden bisher für die Rechnung verwandten Werten 0,342 und 0,322, so erkennt man, daß der Vergleich zwischen den auf Grund der Durchschnittswerte ermittelten Grenzkosten (0,465) und dem Preis für den Frembezug (0,342) zu dem falschen Ergebnis geführt hätte, der Fremdbezug des Dampfes sei kostengünstiger. Es zeigt sich an diesem Beispiel in aller Deutlichkeit, wie wertvoll eine Regressionsanalyse zur Bestimmung der "richtigen" Grenzkosten eingesetzt werden kann, wenn im Rahmen der Kostenrechnung Zahlenmaterial für Optimierungsprobleme bereitgestellt werden muß.

---

1) Die Zahlenwerte wurden übernommen von A. MUIR, a.a.O., S. 100 u. S. 103.

2) Da in die ursprüngliche Regressionsanalyse für die Kosten des Kohleeinsatzes die Mengengröße "Kohleverbrauch" aufgenommen worden war, ist der Mittelwert für diesen Kohleverbrauch (663) mit dem durchschnittlichen Preis für eine Einheit Kohle zu multiplizieren, um die Wertgröße "Kosten" zu erhalten. Vgl. auch S. 133.

5.5 Kostenanalysen im weitesten Sinn

Legt man den Begriff der Kostenanalyse weit aus, so kann man darunter sämtliche Untersuchungen verstehen, die sich in irgendeiner Weise mit dem Zusammenhang zwischen Kosten und ihren Einflußgrößen befassen. Anders formuliert soll die Kostenanalyse "Erkenntnisse über das Kostengefüge und die Kostengebarung vermitteln, die für Kontroll- und Dispositionszwecke verwendbar sind".[1] Somit lassen sich auch die in den vorhergehenden Abschnitten 5.2, 5.3 und 5.4 behandelten Probleme unter dem Oberbegriff der Kostenanalyse subsumieren.

In diesem Abschnitt sollen als Zusammenfassung einige weitere, mehr allgemein gehaltene Aussagen über die Durchführung von Kostenanalysen mit Hilfe der Regressionsrechnung gemacht werden.

Betrachtet man mit MELLEROWICZ als Aufgaben der Kostenanalyse[2]

1. Die Kontrolle der Betriebsgebarung,
2. Die Feststellung von Unwirtschaftlichkeiten,
3. Die Verbesserung der Kostenverrechnung (Kostenstellen- und Kostenträgerrechnung) und
4. Die Vorbereitung der Kostenplanung und der Betriebsdispositionen durch Erforschung der Kostenzusammenhänge und Abhängigkeiten,

so lassen sich sogar sämtliche im Rahmen dieser Arbeit durchgeführten Analysen oder aufgezeigten Anwendungsmöglichkeiten als Kostenanalysen interpretieren.

---

1) K. MELLEROWICZ, a.a.O., II, 2. S. 511.

2) Ebenda. S. 505.

Unterteilt man die allgemeinen Kostenanalysen in zwei Gruppen, die S t r u k t u r a n a l y s e n und die E n t w i c k l u n g s a n a l y s e n,[1] dann kann - wie sich im Verlauf dieser Arbeit gezeigt hat - die Regressionsrechnung insbesondere bei folgenden Aufgaben der Strukturanalyse Anwendung finden:

1. Bei der Feststellung der Faktoren, die die Höhe der Kosten beeinflussen und bei der zahlenmäßigen Ermittlung ihres Einflusses.[2] (Diese Untersuchungen werden auch als A b h ä n g i g k e i t s a n a l y s e n bezeichnet).[3]
2. Bei der Ermittlung von Verteilungsschlüsseln und Zuschlagsgrundlagen für die Gemeinkostenverrechnung (Untersuchung der Verrechenbarkeit von Gemeinkosten auf Kostenstellen und Kostenträger).[4]
3. Bei der Feststellung der Höhe einzelner Kostenarten.[5]

Aber auch im Rahmen der Entwicklungsanalysen bei der Untersuchung der Kostenveränderungen im Zeitablauf und im Soll-Ist-Vergleich hat sich die Regressions- und Korrelationsanalyse als nützliches Hilfsmittel erwiesen.[6] Daneben lassen sich besonders bei der Durchführung von Abhängigkeitsanalysen - einer Vorstufe der Strukturanalysen[7] - deren Aufgabe in der Feststellung der Faktoren liegt, die die Größe und die Entwicklung der Kosten bestimmen, diese statistischen Verfahren vorteilhaft anwenden.

---

1) K. MELLEROWICZ, a.a.O., II, 2, S. 505.

2) Vgl. Abschnitt 5.3, S. 95 ff. und Abschnitt 5.4, S. 118 ff.

3) K. MELLEROWICZ, a.a.O., II, 2. S. 511.

4) Vgl. Abschnitt 4.4, S. 58 ff.

5) Vgl. Abschnitt 4.2, S. 37 ff.

6) Vgl. S. 81 ff., insbesondere aber den Abschnitt 6.2, S. 168 ff.

7) K. MELLEROWICZ, a.a.O., II, 2, S. 527.

Im Rahmen der Planung von Gemeinkosten wurde dargelegt, daß sich die Regressionsanalyse aus verschiedenen Gründen nicht gut zur Ermittlung der Sollkostenkurve eignet,[1] und daß sog. "technische" Untersuchungen in der Regel empirisch-statistischen vorzuziehen sind. Dabei wurde im Abschnitt 5.2 die Frage offen gelassen, wie diese technischen oder ingenieurmäßigen Kostenstudien in der Praxis aussehen. Hierauf soll im folgenden kurz eingegangen werden, da auch in diesem Fall eine Regressionsanalyse durchaus wichtige Erkenntnisse vermitteln kann.

Als Beispiel für eine solche Analyse lassen sich besonders gut die Energiekosten - bei Außerachtlassung der Preiskomponente - auch die im Produktionsprozeß eingesetzten Mengen an Energie heranziehen, wie zum Beispiel Kohle-, Heizöl- oder Stromverbrauch. Für diese Untersuchung der Abhängigkeit des Verbrauchs eines Energieträgers von mehreren verschiedenen Einflußfaktoren bietet sich die m u l t i p l e Regressionsanalyse geradezu an.[2] So lassen sich zum Beispiel die Stromkosten beim Walzen in einem Walzwerk in Abhängigkeit von den Einflußgrößen Walzzeit und Walzmenge untersuchen, wobei die Walzzeit wiederum von dem Ausgangsquerschnitt vor dem Walzen, dem Querschnitt des Walzproduktes, der Art des zu walzenden Materials und der Glühtemperatur

---

1) Vgl. S. 81 ff., insbesondere S. 86 ff., wo ausführlich auf die Problematik der Gewinnung von Plankosten aus Istkosten der Vergangenheit mit Hilfe der einfachen linearen Regressionsanalyse eingegangen wird.

2) Vgl. hierzu auch das Beispiel "Koksverbrauch bei der Thomasroheisenerzeugung in Abhängigkeit von verschiedenen Einflußgrößen", S. 107ff., das von MATT als schrittweise Regressionsanalyse durchgeführt wurde. Vgl. auch G. MATT: Die schrittweise Regressionsanalyse ..., a.a.O., S. 261 ff.

abhängt.[1] Diese sechs Einflußgrößen lassen sich als unabhängige Variable einer linearen multiplen Regression auffassen, so daß anhand der in der Vergangenheit beobachteten Werte sich die Veränderungen im Stromverbrauch auf Grund von Veränderungen in den beeinflussenden Faktoren durch die Regressionsgleichung zahlenmäßig ermitteln lassen. So gesehen läßt sich die Regressionsanalyse als eine Methode zur Aufstellung von Verbrauchsfunktionen bezeichnen.[2] Es wurde ausführlich dargestellt, welche Vorteile gerade die schrittweise Regressionsanalyse in diesem Zusammenhang bietet.[3]

Zusammenfassend kann festgestellt werden, daß die Regressionsanalyse praktisch auf dem "Umweg" über die hier gezeigte Form der Kostenanalyse einen wichtigen Beitrag zur "technischen" oder "synthetischen" Planung der Kosten - auch für die Konstruktion der Sollkostenkurve im Rahmen der flexiblen Plankostenrechnung - leisten kann. Gerade eine multiple Regressionsgleichung, in der auf Grund des Vorgehens der schrittweisen Analyse alle wichtigen (signifikanten) Einflußgrößen enthalten sind, ermöglicht eine gute Anpassung an eventuelle zukünftige Veränderungen in den beeinflussenden Faktoren.

---

1) Vgl. F. HENZEL: Kosten und Leistung. 3. unveränd. Aufl. der Kostenanalyse. Stuttgart 1957. S. 143., wo als weiteres Beispiel der Kohleverbrauch von Glüh- und Anwärmeöfen in Abhängigkeit von Temperatur und Brenndauer des leeren Ofens, der Menge des eingesetzten Materials, der Unterbrechungsdauer zwischen den Glühvorgängen und der Glühtemperatur des eingesetzten Materials untersucht wird. Vgl. S. 149 ff.

2) Vgl. auch H.-R. HABERBECK: Zur Beschreibung der Abhängigkeitsstruktur des Produktionsfaktorverbrauches. "Zeitschrift für Betriebswirtschaft". 38 (1968) 12, S. 905 ff., hier S. 913 ff.

3) Vgl. S. 99 ff.

In den Rahmen der Kostenanalyse gehören weiterhin Kostenuntersuchungen, die der Preisfestsetzung dienen. Solche Analysen wurden zum Beispiel in den Vereinigten Staaten von MC CLENON durchgeführt, der in einer kleineren Reparaturwerkstatt für Büromaschinen die direkten Lohnkosten[1] für die verschiedenen zu reparierenden Maschinenarten mit Hilfe der multiplen Regressionsanalyse zu finden versuchte,[2] wobei die Lohnkosten der verschiedenen Reparaturarbeiten die Grundlage für den Pauschalpreis für die Reparatur jeder Maschine zuzüglich des benötigten Materials bilden sollten. Es handelt sich bei dieser Untersuchung demnach um eine Schätzung der Stückkosten aus den aufgezeichneten Gesamtkosten der vergangenen Rechnungsperioden, wobei die partiellen Regressionskoeffizienten die jeweils für eine Reparatur eines bestimmten Maschinentyps im Durchschnitt anfallenden Lohnkosten (bzw. Arbeitszeiten, wenn man den Stundenlohn unberücksichtigt läßt) angeben.

Man mag einwenden, daß einem größeren Unternehmen mit ausgebautem Rechnungswesen in aller Regel die pro Reparatur einer Schreib-, Rechen- oder Additionsmaschine im Durchschnitt anfallenden Lohnkosten auch ohne Regressionsanalyse bekannt sein werden. Für kleinere Unternehmen jedoch, auf die MC CLENON in seinem Beispiel verweist,[3] erscheint das beschriebene Verfahren der multiplen Regressionsanalyse geeignet, und zwar besonders in Situationen, in denen schnelle Entscheidungen, zum Beispiel über die Annahme oder Ablehnung eines Auftrags oder über Selbstherstellung oder Fremdbezug,[4] getroffen werden müssen. Häufig wird in die-

---

1) Im Original: "direct labor costs".

2) P.R. MC CLENON: Cost Finding Through Multiple Correlation Analysis. "The Accounting Review". 38 (1963) 3, S. 540 ff.

3) Ebenda. S. 540.

4) Wenn zum Beispiel ein neues Angebot eines Lieferanten abgegeben wird, das günstigere als die bisherigen Einkaufsbedingungen enthält.

sen Betrieben die Kostenrechnung nicht so exakt durchgeführt, daß Unterlagen für diese Entscheidungen sofort, d.h. ohne weitere Analysen durchführen zu müssen, zur Verfügung stehen.[1] Allerdings wird man sich auf Regressionsanalysen mit drei oder höchstens vier unabhängigen Variablen beschränken müssen, da sonst der Rechenaufwand zu umfangreich wird, wenn - was in diesem Fall wohl unterstellt werden muß - keine programmierbaren Rechenmaschinen zur Durchführung der Analyse zur Verfügung stehen.[2]

5.6 Zuverlässigkeit der Schätzungen - Bildung von Konfidenzintervallen

5.6.1 Problemstellung

In den vorhergehenden Abschnitten dieses 5. Kapitels wurde anhand von Beispielen gezeigt, auf welche Weise die Regressionsanalyse dazu dienen kann, aus vorhandenem Datenmaterial S c h ä t z w e r t e für Kosten oder Verbrauchsmengen zu ermitteln. Dabei wurde bereits an mehreren Stellen darauf hingewiesen, daß diese Regreßwerte den Charakter von M i t t e l w e r t e n, D u r c h s c h n i t t s- oder E r w a r t u n g s w e r t e n haben, die nicht den a l l e i n r i c h t i g e n Kosten- oder Verbrauchswert

1) So hatte der von MC CLENON untersuchte Betrieb lediglich e i n e Berechnungsgrundlage für die Überholung von Schreib-, Additions- und Rechenmaschinen, obwohl bekannt war, daß die Durchsicht der verschiedenen Maschinenarten in seiner Höhe unterschiedlichen Arbeitsaufwand erforderte.

2) Zur Frage des Rechenaufwands bei manueller oder elektronischer Ermittlung der Regressionsparameter vgl. 9. Anhang I: Rechnerische Durchführung von Regressions- und Korrelationsanalysen, S. 252 ff., hier insbesondere S. 257, S. 263 u. S. 265.

wiedergeben können.[1] Das ist nicht möglich, weil "sich eine Erscheinung auf Grund der vielfältigen Wirklichkeit in vielen unterschiedlichen Einzelwerten äußern kann [und] die empirischen Einzelwerte um den Durchschnittswert streuen".[2] Die Anwendung der Regressionsanalyse verlangt also ein "Denken in Toleranzen",[3] wobei diese Toleranzen (Konfidenzintervalle, Vertrauensbereiche) unter Heranziehung der Grundlagen der Wahrscheinlichkeitstheorie berechnet werden können. Das bedeutet, daß die Möglichkeit besteht, für jeden mit Hilfe der Regressions- oder Korrelationsanalyse ermittelten Schätzwert (Gesamtkosten, Grenzkosten, Verbrauchsmengen etc.) einen Bereich anzugeben, in dem diese Werte mit einer bestimmten statistischen Wahrscheinlichkeit erwartet werden können.[4]

In den folgenden Abschnitten soll gezeigt werden, welche statistischen Größen für die Berechnung dieser Konfidenzintervalle herangezogen werden können, wie sie ermittelt werden und auf welchen Voraussetzungen ihre Anwendung beruht.

5.6.2 Erläuterung des Zufallsmodells

Bei der Überprüfung unserer Kosten- und Verbrauchsschätzungen mit Hilfe der Wahrscheinlichkeitstheorie durch die Bil-

---

1) Vgl. auch den Abschnitt 6.2.

2) E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 103.

3) Ebenda. S. 106.

4) Vgl. hierzu auch G. FRAAS: Neuere Aspekte des Kostenrechnungsprozesses. "Statistische Praxis. Zeitschrift für Statistik und Rechnungswesen". 20 (1965) 10, S. 407 ff., hier S. 408, der auf die Bedeutung des Übergangs von der kausalen zur wahrscheinlichkeitsstatistischen Betrachtungsweise in der Kostenrechnung ausführlich eingeht.

dung von Konfidenzintervallen stellt sich zunächst die Frage, ob der von uns vorgenommenen Anwendung der Regressions- und Korrelationsanalyse auf kostenrechnerische Probleme überhaupt ein Zufallsmodell zugrundeliegt. Nur wenn diese Frage bejaht werden kann, lassen sich mit Hilfe der Wahrscheinlichkeitsrechnung Aussagen über die Zuverlässigkeit unserer Kostenschätzungen treffen.

Es ist also zunächst zu prüfen, ob die untersuchten Kosten oder Kostenabweichungen Z u f a l l s v a r i a b l e im Sinn der Wahrscheinlichkeitsrechnung sind. Unter einer zufälligen oder stochastischen Variablen versteht man eine solche Variable, "die verschiedene Werte mit den ihnen zukommenden Wahrscheinlichkeiten anzunehmen vermag".[1] Von einer Zufallsvariablen kann man also nicht von vornherein sagen, welchen speziellen Wert sie in einem bestimmten Fall annehmen wird, da man den einzelnen Werten - als den Ergebnismöglichkeiten des Zufallsexperiments - lediglich bestimmte Wahrscheinlichkeiten zuordnen kann. Anders formuliert: Bei einer Wiederholung des Experiments läßt sich das Ergebnis einer konkreten Realisation nicht mit Sicherheit vorhersagen.

Im Abschnitt 2.1 des 2. Kapitels[2] wurde bereits ausführlich darauf eingegangen, daß die Aufgaben der Regressions- und Korrelationsanalyse vor allem darin zu sehen sind, stochastische Zusammenhänge zwischen Variablen zu untersuchen. Eine solche stochastische, d.h. wahrscheinlichkeitstheoretische Abhängigkeit besteht auch zwischen den Kosten oder Verbrauchsmengen und ihren Einflußfaktoren. Bei einem gegebenen Wert der unabhängigen Veränderlichen, d.h. bei einer bestimmten Produktionsmenge, lassen sich die Kosten

---

1) O. ANDERSON, a.a.O., S. 91.

2) Vgl. S. 9 f.

nicht eindeutig in ihrer Höhe angeben, sondern streuen z u f a l l s b e d i n g t in einem Intervall. Gerade auf Grund dieser Tatsache erweist sich die Anwendung der Regressionsanalyse auf kostenrechnerische Probleme als wertvolles Hilfsmittel für die Planung, Verrechnung und Kontrolle von Kosten.

Wir können also die Kosten (Verbrauchsmengen) und ihre Einflußgrößen, die "Realisierungen aus einer Menge von theoretisch unendlich vielen Möglichkeiten"[1] darstellen, als zufällige Veränderliche im Sinn der Wahrscheinlichkeitsrechnung ansehen.

5.6.3 Verteilung der Zufallsvariablen

Interpretieren wir die Kosten und ihre Einflußgrößen als zufällige Veränderliche, dann läßt sich jedem Wert eine bestimmte Wahrscheinlichkeit zuordnen. Stellt man diese Wahrscheinlichkeiten beginnend mit dem kleinsten Wert der Zufallsvariablen nebeneinander, so erhält man die W a h r s c h e i n l i c h k e i t s v e r t e i l u n g der Zufallsvariablen.

Da wir die Zahlenwerte (Kosten, Verbrauchsmengen) auf Grund einer empirischen Stichprobe aus einer übergeordneten Gesamtheit erhalten haben, bezeichnen wir ihre Verteilung als e m p i r i s c h e Wahrscheinlichkeitsverteilung. Wir können zunächst über ihr Verteilungsgesetz keine allgemeinen Aussagen machen. Jedoch wird die rechnerische Behandlung solcher empirischer Verteilungen durch die Tatsache

---

1) G. FRAAS, a.a.O., S. 408.

erleichtert, daß sie sich häufig recht gut durch theoretische Wahrscheinlichkeitsverteilungen approximieren, d.h. näherungsweise beschreiben lassen.

Die für die Praxis bedeutendste und am häufigsten verwendete theoretische Verteilung ist die sog. "N o r m a l v e r t e i l u n g", da viele empirische Verteilungen gerade durch sie gut approximiert werden können. Auf ihre Grundlagen soll im folgenden Abschnitt kurz eingegangen werden.

5.6.3.1 Die Normalverteilung

Die Normalverteilung - auch als GAUSS-LAPLACE-Verteilung bezeichnet - ist eine stetige zweiparametrische Verteilung mit dem Erwartungswert (Mittelwert) E (X) = μ und der Varianz (Streuung) V (X) = σ^2, wobei der Parameter μ die Lage und Parameter σ^2 die Streuung der Verteilung charakterisiert.[1] Durch Standardisierung dieser Normalverteilung läßt sich jede normalverteilte Zufallsvariable X in eine Zufallsvariable Z überführen, mit E (Z) = μ = 0 als Mittelwert und V (Z) = σ^2 = 1 als Varianz.[2] Hierdurch läßt sich eine Tabelle der Standardnormalverteilung aufstellen, in der sich über die Ermittlung einzelner z-Werte bei gegebenem Mittelwert und gegebener Standardabweichung die Wahrscheinlichkeit für das Auftreten bestimmter vom Mittelwert abweichender Werte ablesen läßt,[3] oder anders formuliert, man kann angeben, "mit welcher Wahrscheinlichkeit ein standardisierter Wert der normalverteilten Veränderli-

1) Zur ausführlichen Darstellung der Normalverteilung vgl. jedes Lehrbuch der Statistik, z.B. O. ANDERSON, a.a.O., S. 249 u. E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 219 ff.

2) Vgl. auch den Abschnitt 6.2.2.

3) Vgl. die Tabellen 1* u. 2* der Normalverteilung im Anhang II. S. 284 f.

chen in ein vorher gewähltes Intervall fallen wird".[1]

Bei der Überprüfung unserer Kostenschätzungen stellt sich nun die Frage, ob eine Anwendung dieser Normalverteilung zulässig ist oder ob die Kosten oder Kostenabweichungen einem anderen Verteilungsgesetz unterliegen.

Zur Beantwortung dieser Frage können wir von den folgenden Überlegungen ausgehen. Im Kostenverursachungsprozeß treten sowohl z u f ä l l i g e als auch s y s t e m a t i s c h e Fehler auf.[2] Systematische oder "beherrschbare" Fehler entstehen zum Beispiel durch falsche oder ungenaue innerbetriebliche Leistungsverrechnung oder Zurechnung von Gemeinkosten auf Kostenstellen und Kostenträger, weil hierdurch die richtige Beziehung zwischen den Kosten und ihren Einflußfaktoren verschleiert oder verzerrt wird.[3] Liegen solche Fehler in größerem Umfang im Rechnungswesen vor, kann durch die eingetretenen Verzerrungen die Annahme der Normalverteilung der Kosten nicht aufrecht erhalten werden. Gelingt es jedoch, diese Fehler zu eliminieren, dann kann man von folgender Annahme ausgehen:
"Solange nicht empirische Versuche, die außerordentlich zeitraubend wären, weil der Erfassungs- und Abrechnungsprozeß nicht darauf eingerichtet ist, diese Hypothese umstoßen, kann angenommen werden, daß ein N o r m a l v e r t e i l u n g s g e s e t z z u g r u n d e l i e g t[4]".[5]
Dieser Auffassung ist auch SCHÖNFELD, dem es ebenfalls vertretbar erscheint, mit der Hypothese einer Normalverteilung

---

1) E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 221.

2) G. FRAAS, a.a.O., S. 408. Hier werden weitere Beispiele für systematische und auch für zufällige Fehler aufgeführt. Vgl. in diesem Zusammenhang auch den Abschnitt 6.2, S. 168 ff.

3) Vgl. hierzu den Abschnitt 4.4, S. 58 ff.

4) Vom Verfasser gesperrt.

5) G. FRAAS, a.a.O., S. 408.

zu arbeiten, bevor nicht anhand praktischer Untersuchungen nachgewiesen worden ist, daß diese Unterstellung nicht haltbar ist.[1] Sind die Zufallsvariablen aber normalverteilt, dann unterliegen auch die Regressionsparameter dem Normalverteilungsgesetz. Wir wollen uns dieser Argumentation anschließen, jedoch noch einmal betonen, daß die Gültigkeit der Normalverteilung eine Arbeitshypothese bleiben muß, solange nicht bewiesen worden ist, daß die untersuchten Kosten und ihre Einflußfaktoren tatsächlich diesem Gesetz unterliegen.

5.6.3.2 Der zentrale Grenzwertsatz

Allerdings läßt sich unter Umständen auch dann mit einer Normalverteilung arbeiten, wenn die Zufallsvariablen selbst nicht normalverteilt sind. Der Beweis für diese Aussage läßt sich mit Hilfe des "zentralen Grenzwertsatzes" führen.[2] Dieses Theorem besagt - in einer für die praktische Stichprobenerhebung angemessenen Fassung:[3]

> "Wenn man Stichproben jeweils vom Umfang n einer und derselben Gesamtheit höherer Ordnung nach einem Zufallsmodell entnimmt, so wird - bei einem praktisch beinahe beliebigen Verteilungsgesetz derselben - das Verteilungsgesetz der arithmetischen Durchschnitte dieser Stichproben (und folglich auch das der ihnen zugrunde liegenden Summen) mit zunehmendem Umfang n stochastisch gegen das normale Verteilungsgesetz konvergieren."

Stark vereinfacht, aber für unsere Zwecke noch genügend exakt, formuliert sagt der zentrale Grenzwertsatz aus, daß "die Mittelwerte der Stichprobe annähernd normalverteilt

---

1) H.-M. SCHÖNFELD: Die Anwendung ..., a.a.O., (1968), S. 248.

2) Vgl. auch A. HALD, a.a.O., S. 189.

3) O. ANDERSON, a.a.O., S. 136 f.

sind, falls die Stichprobe nicht zu klein ist".[1)]

Wir können damit bei den folgenden Untersuchungen von den Voraussetzungen der Normalverteilung ausgehen, wenn auch ihre Existenz lediglich wahrscheinlich ist und nicht bewiesen werden konnte.

Aber auch wenn sich in einem bestimmten Fall auf Grund des beobachteten Datenmaterials und eines geringen Stichprobenumfangs die Annahme eines Normalverteilungsgesetzes als unhaltbar erweisen sollte, lassen sich Konfidenzintervalle für die regressionsanalytischen Schätzwerte ermitteln. Hierauf soll der Vollständigkeit halber im Abschnitt 5.6.3.3 kurz eingegangen werden, auch wenn bei den an späterer Stelle durchgeführten Untersuchungen stets eine Normalverteilung (bzw. Studentverteilung) zugrundegelegt wird.

5.6.3.3 Die Ungleichung von TSCHEBYSCHEFF

Selbst bei unbekanntem Verteilungsgesetz der Schätzwerte und Regressionsparameter bietet die Statistik die Möglichkeit, Vertrauensgrenzen mit Hilfe der sog. "Ungleichung von TSCHEBYSCHEFF" zu ermitteln.[2)]

Dieses Theorem sagt aus, daß ohne Rücksicht auf die Häufigkeitsverteilung von Variablen und ohne Rücksicht auf ir-

---

1) H.-M. SCHÖNFELD: Die Anwendung statistischer Methoden in der Kostenrechnung (II). "Kostenrechnungspraxis. Zeitschrift für Betriebsabrechnung, Kostenrechnung und -planung". O.J. (1969) 1, S. 19 ff., hier S. 20 Fußnote 1. - Auf die Problematik der Gewinnung von relativ großen Stichproben bei Kostenuntersuchungen wird im Abschnitt 7.3 eingegangen.

2) Auch als "Ungleichung von BIENAYME-TSCHEBYSCHEFF" bezeichnet. Vgl. O. ANDERSON, a.a.O., S. 131.

gendwelche Beziehungen funktioneller, kausaler oder anderer Natur, die zwischen ihnen existieren mögen,[1] stets eine bestimmte Wahrscheinlichkeit für die Abweichungen eines Variablenwertes vom Mittelwert angegeben werden kann. Dabei ist die Wahrscheinlichkeit, daß eine Variable um mehr als das k-fache ihrer Standardabweichung von ihrem Mittelwert abweicht $\leq \frac{1}{k^2}$.

Wählen wir zum Beispiel für k den Wert k = 2, was besagt, daß die Grenzen der Konfidenzintervalle im Abstand von zwei Standardabweichungen um den Mittelwert gezogen werden sollen, so beträgt die Wahrscheinlichkeit dafür, daß Beobachtungswerte außerhalb dieser Grenzen fallen, weniger als $\frac{1}{4}$ oder 25 %. Anders formuliert, die Wahrscheinlichkeit, daß zufällige Beobachtungswerte oder die in unseren Beispielen häufig untersuchten Kostenabweichungen innerhalb dieses Bereichs von zwei Standardabweichungen um den Mittelwert liegen, ist größer als 75 %. Diese Aussage läßt sich für andere Vielfache der Standardabweichung entsprechend treffen. So fallen zum Beispiel weniger als 11,1 % aller Werte außerhalb des Bereichs $\mu \pm 3\sigma$, denn es gilt:

$$\frac{1}{k^2} = \frac{1}{3^2} = 0{,}111.$$

Es leuchtet ein, daß bei unbekanntem Verteilungsgesetz die Aussagen über die Grenzen der Konfidenzintervalle nicht mit der gleichen Exaktheit getroffen werden können, die bei Anwendung der Normalverteilung möglich ist. Dennoch kann auch eine ungenauere Aussage sicherlich in vielen Situationen der betrieblichen Praxis noch zu a u s r e i c h e n d e x a k t e n Ergebnissen führen und damit eine richtige Entscheidung erlauben.[2]

---

1) O. ANDERSON, a.a.O., S. 132.; Vgl. auch F.E. CROXTON u. D.J. COWDEN, a.a.O., S. 221.

2) Zu weiteren nichtparametrischen Testverfahren vgl. z.B. F.R. OLIVER: A Cross-section Study of Marginal Cost. "Applied Statistics. A Journal of the Royal Statistical Society". 11 (1962) 2, S. 69 ff., hier S. 76 f. Zur mathematischen Formulierung und Programmierung von Testverfahren im verteilungsfreien Fall mit Hilfe der Un-

5.6.4 Standardfehler der Regressionskoeffizienten

Zunächst soll anhand eines Beispiels erläutert werden, wie sich die Standardfehler der Regressionskoeffizienten ermitteln lassen. Die Tatsache, daß die mit Hilfe der Regressionsfunktion errechneten Werte ihrem Charakter nach Mittelwerte sind, die in bestimmten Grenzen streuen, gilt für die Regressionskonstante a_{yx} und den Regressionskoeffizienten b_{yx} (im Fall der einfachen Regression) und die Regressionskonstante $a_{y.12...m}$ und die partiellen Regressionskoeffizienten $b_{y1.23...m}$, $b_{y2.13...m}$ ... $b_{ym.12...m-1}$ (im Fall der multiplen Regression.) Diese aus einer Stichprobe ermittelten Regressionskoeffizienten, die erwartungstreue und konsistente Schätzwerte für die Regressionskoeffizienten der Grundgesamtheit darstellen,[1] streuen mit der Standardabweichung σ_b um den wahren Regressionskoeffizienten β der Grundgesamtheit und sind - wie wir annehmen wollen, normalverteilt.[2] Da die Standardabweichung des Regressionskoeffizienten - wir gehen zunächst von einer einfachen Regression aus - in der Grundgesamtheit ($\sigma_{b_{yx}}$) in den meisten Fällen unbekannt ist, muß sie auf Grund des Stichprobenzahlenmaterials nach der Formel[3]

$$s_{b_{yx}} = \frac{s_{y.x}}{s_x\sqrt{n}} \qquad (5.6)$$

Fortsetzung der Fußnote 2) von S. 149:
gleichung von TSCHEBYSCHEFF und ihrer Anwendung auf regressionsanalytische Schätzwerte vgl. G. LUDA: Programm zur Durchführung von Regressionsanalysen auf elektronischen Datenverarbeitungsanlagen. "Allgemeines Statistisches Archiv". 46 (1962) 2, S. 122 ff.

1) Vgl. z.B. auch E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 226.; A. HALD, a.a.O., S. 208 f. Auf die Bedeutung der Begriffe "konsistent" und "erwartungstreu" kann hier nicht näher eingegangen werden. Vgl. dazu jedes Lehrbuch der Statistik.

2) Vgl. S. 145 ff. u. E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 225.

3) Zu den Formeln vgl. E. FÖRSTER u. F. EGERMAYER, a.a.O., S.223.; M. EZEKIEL u. K.A. FOX, a.a.O., S. 281.

ermittelt werden. In dieser Formel bezeichnet $s_{b_{yx}}$ den Standardfehler des Regressionskoeffizienten, $s_{y \cdot x}$ den Standardfehler der Schätzung[1] und s_x die Standardabweichung der x-Werte.[2] Der Stichprobenumfang wird durch n angegeben. Um hieraus einen unverzerrten Schätzwert für $\hat{\sigma}_{b_{yx}}$ zu erhalten,[3] sind für die Größen $s_{y \cdot x}$ und s_x ebenfalls die unverzerrten Schätzwerte $\hat{\sigma}_{y \cdot x}$ und $\hat{\sigma}_x$ einzusetzen,[4] so daß man folgenden Ausdruck erhält:

$$\hat{\sigma}_{b_{yx}} = \frac{\hat{\sigma}_{y \cdot x}}{\hat{\sigma}_x \sqrt{n}}, \qquad (5.7)$$

oder in ausführlicher Schreibweise:

$$\hat{\sigma}_{b_{yx}} = \frac{s_{y \cdot x} \sqrt{\frac{n}{n-2}}}{s_x \sqrt{\frac{n}{n-1}} \sqrt{n}} = \frac{s_{y \cdot x}}{s_x \sqrt{n}} \sqrt{\frac{n-1}{n-2}}. \qquad (5.8)^{5}$$

Damit läßt sich der Standardfehler des Regressionskoeffizienten aus dem Standardfehler der Residuen[6] und der Stan-

---

1) Zur ausführlichen Darstellung des Standardfehlers der Schätzung (Standardabweichung der Residuen) vgl. den Abschnitt 5.6.5, S. 155 ff.

2) Die Standardabweichung der x-Werte ist definiert als

$$s_x = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}.$$

3) Vgl. auch F.E. CROXTON u. D.J. COWDEN, a.a.O., S. 644.

4) Vgl. auch den Anhang zur Symbolik.

5) Vgl. auch E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 226.

6) Vgl. auch den Abschnitt 6.2.

dardabweichung der x-Werte unter Berücksichtigung des Stichprobenumfangs ermitteln. Bei multiplen Regressionen ergeben sich die Standardfehler der partiellen Regressionskoeffizienten in ganz ähnlicher Weise, wenn auch überwiegend in der Literatur etwas andere Formeln angegeben werden.[1)]

Wir wollen die praktische Verwendbarkeit dieses Streuungsmaßes noch durch ein Zahlenbeispiel erläutern. Im Abschnitt 5.4.2.2 wurde anhand einer Stichprobe vom Umfang n = 10 dargestellt, wie sich die Grenzkosten mit Hilfe von partiellen Regressionskoeffizienten schätzen lassen. Dabei hatten sich für die drei in die Analyse aufgenommenen Variablen x_1, x_2 und x_3 folgende Koeffizienten, die mit den Grenzkosten der drei Produkte identisch sind, ergeben:[2)]

$$b_{y1.23} = 48{,}67,$$
$$b_{y2.13} = 193{,}94 \quad \text{und}$$
$$b_{y3.12} = 100{,}81.$$

Wir wollen nunmehr prüfen, wie zuverlässig diese auf Grund einer so kleinen Stichprobe gewonnenen Werte die wahren Grenzkosten in der Grundgesamtheit wiedergeben. Dazu ermitteln wir ihre zugehörigen Standardfehler und erhalten:[3)]

$$\hat{\sigma}_{b_{y1.23}} = 31{,}69,$$
$$\hat{\sigma}_{b_{y2.13}} = 22{,}20 \quad \text{und}$$
$$\hat{\sigma}_{b_{y3.12}} = 12{,}75.$$

---

1) Zur ausführlichen Darstellung vgl. E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 223.; M. EZEKIEL u. K.A. FOX, a.a.O., S. 283 ff.

2) Vgl. S. 127 f.

3) Entnommen aus: J.S. CHIU u. D.T. DE COSTER, a.a.O., S. 676.

Wenn wir unsere Aussagen über den wahren Grenzkostenbetrag jeder Produktart in der Grundgesamtheit mit einem Sicherheitsgrad von 95 % treffen wollen, erhalten wir aus der Tabelle der Studentverteilung[1] einen z-Wert von $z_{0,95} = 2,45$ bei 6 Freiheitsgraden.[2]

Für das Konfidenzintervall für die wahren Grenzkosten des Produktes 1 gilt damit folgende Beziehung:

$$b_{y1.23} - 2,45\,\hat{\sigma}_{b_{y1.23}} < \beta_{y1.23} \leq b_{y1.23} + 2,45\,\hat{\sigma}_{b_{y1.23}},$$

$$48,67 - 2,45 \cdot 31,69 < \beta_{y1.23} \leq 48,67 + 2,45 \cdot 31,69,$$

$$-28,99 < \beta_{y1.23} \leq 126,33.$$

Für die Grenzkosten der Produkte 2 und 3 ergeben sich durch die analoge Berechnungsweise - bei gleichem Signifikanzniveau - folgende Konfidenzintervalle:

Produkt 2: $139,55 < \beta_{y2.13} \leq 248,33,$

Produkt 3: $69,57 < \beta_{y3.12} \leq 132,05.$

Insbesondere für die Grenzkosten des Produktes 1 zeigt sich eine sehr große Streuung, die sogar negative Werte umfaßt, ein Ergebnis, das vom kostenrechnerischen Standpunkt aus

---

1) Die Normalverteilung kann hier wegen der unbekannten Werte für die Standardfehler der Regressionskoeffizienten der Grundgesamtheit und des kleinen Stichprobenumfangs von n = 10 nicht verwendet werden. Vgl. dazu jedes Lehrbuch der Statistik, in dem Häufigkeitsverteilungen behandelt werden.

2) Die Freiheitsgrade errechnen sich folgendermaßen:
$\nu = n - k - 1 = 10 - 3 - 1 = 6.$
ν = Zahl der Freiheitsgrade, k = Zahl der in die Regression aufgenommenen unabhängigen Variablen.

nur als sinnlos bezeichnet werden kann und das zu der Feststellung führen muß, daß die Grenzkosten-Schätzwerte nur mit Vorsicht verwendet werden dürfen.

Anstelle der Bestimmung eines Konfidenzintervalls für die Regressionskoeffizienten kann man auch über die Bestimmung von z-Werten eine Hypothesenprüfung vornehmen. Hierbei setzt man einen z-Wert als kritischen Wert fest und sieht die Regressionskoeffizienten nur dann als statistisch gesichert an, wenn der auf Grund des Stichprobenergebnisses errechnete z-Wert den kritischen Wert überschreitet. Wie eine solche Hypothesenprüfung im einzelnen durchgeführt wird, wird im Abschnitt 6.2 dargestellt. Deshalb kann an dieser Stelle der Hinweis auf diese Möglichkeit genügen.[1)]

Die relativ große Streuung aller Regressionskoeffizienten resultiert aus dem geringen Stichprobenumfang von n = 10, kann jedoch durch Einbeziehung weiterer Werte in die Analyse, d.h. durch Erhöhung des Stichprobenumfangs, erheblich verringert werden.[2)] Sie sollte darum kein Argument für die Ablehnung dieser Berechnungsmethode bilden, wenn in der Praxis Zahlenwerte aus genügend vielen Beobachtungsperioden zur Verfügung stehen. Von der mathematischen Lösungsmöglichkeit des Gleichungssystems aus gesehen lassen sich multiple Regressionsanalysen durchführen, wenn die Zahl der in die Stichprobe einbezogenen Beobachtungswerte um eins größer ist als die Anzahl der unabhängigen Variablen.[3)] Allerdings sagt dieser Satz nichts aus über die Zuverlässigkeit solcher Schätzungen, die jedoch in der Praxis

---

1) Vgl. auch T. YAMANE, a.a.O., S. 390 ff.

2) Bei einer Erhöhung des Stichprobenumfangs vermindern sich die Standardfehler der Regressionskoeffizienten etwa um die Quadratwurzel aus der Zahl der zusätzlich in die Analyse aufgenommenen Beobachtungswerte.

3) Vgl. G.J. BENSTON, a.a.O., S. 663.

eine entscheidende Rolle spielt. Um zu vernünftigen, d.h. im üblichen Rahmen statistisch gesicherten Ergebnissen zu gelangen, sind selbstverständlich wesentlich mehr Beobachtungswerte erforderlich. Auch der dann zunehmende Rechenaufwand dürfte durch den Einsatz von Datenverarbeitungsanlagen leicht zu bewältigen sein, zumal lediglich die Beobachtungsreihen "verlängert", aber keine neuen unabhängigen Variablen in die Rechnung aufgenommen werden.[1)]

5.6.5 Standardfehler der Schätzung, Standardfehler der Regressionsgeraden und Standardfehler einer individuellen Prognose

In gleicher Weise, wie wir für die Regressionskoeffizienten Vertrauensgrenzen bestimmt haben, können wir auch für die Schätzwerte selbst (y'_i - Werte) Konfidenzintervalle ermitteln. Auch hier gilt die Aussage, daß die Regreßwerte (Kosten- oder Verbrauchsschätzungen) Stichprobenmittelwerte sind, die gewissen Schwankungen unterliegen, die mit einer bestimmten statistischen Wahrscheinlichkeit angegeben werden können.[2)]

Diese Bildung von Vertrauensgrenzen für die Regressionslinie kann über die Ermittlung des Standardfehlers $6_{y.x}$ geschehen.[3)] Diese Größe, die auch als Standardabweichung der Residuen bezeichnet wird, läßt sich aus der Stichprobe

---

1) Auf die Berechnung von Konfidenzintervallen für die Regressions k o n s t a n t e "a" soll hier verzichtet werden. Sie ist in ähnlicher Weise möglich, wie das für die Regressionskoeffizienten gezeigt wurde. Vgl. z.B. N.R. DRAPER u. H. SMITH, a.a.O., S. 21.

2) E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 229.

3) Vgl. zur ausführlichen Darstellung dieses Verfahrens den Abschnitt 6.2.

wie folgt errechnen:[1]

$$s_{y.x} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - y_i')^2}{n - 2}} \quad . \qquad (5.9)$$

In dieser Gleichung werden - wie in den vorhergehenden Abschnitten dieser Arbeit - die beobachteten Kosten bzw. Verbrauchsmengen mit y_i und die durch die Regressionsfunktion geschätzten Werte mit y_i' bezeichnet, n gibt den Stichprobenumfang wieder.

Der Standardfehler der Schätzung $s_{y.x}$ ist ein Maß für die Streuung der Residuen um die Regressionsgerade, d.h. ein Maß für die nicht durch die Regression erklärten Veränderungen in der abhängigen Veränderlichen y. Die im Abstand eines bestimmten Vielfachen von $\hat{\sigma}_{y.x}$ um die Regressionslinie gezogenen Grenzen ergeben dann P a r a l l e l e n zur Regressionsgeraden.[2] Dieses Verfahren hat den Vorteil der einfachen Berechnungsmöglichkeit für die Konfidenzintervalle, da unabhängig von dem Wert der Variablen x immer dieselbe Standardabweichung zugrunde gelegt wird (aus diesem Grund bilden die Vertrauensgrenzen P a r a l l e l e n zur Regressionsgeraden).

Ein exakteres Verfahren zur Bestimmung von Konfidenzintervallen um die Regressionslinie geht von der Überlegung aus,

---

1) Diese Formel gilt für die einfache Regression. Vgl. z.B. E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 199. Um zu einem unverzerrten Schätzwert für den Standardfehler der Schätzung in der Grundgesamtheit ($\sigma_{y.x}$) zu gelangen, ist $s_{y.x}$ mit einem Korrekturfaktor zu berichtigen. Es gilt dann folgende Beziehung:

$$\hat{\sigma}_{y.x} = s_{y.x} \sqrt{\frac{n}{n - 2}} \quad .$$

2) Vgl. Abb. 5.4., S. 158.

daß jeder Regreßwert y', der zu einem bestimmten x-Wert gehört, in gewissen Grenzen streuen kann. Unter Berücksichtigung dieser Streuung gelangt man zu der folgenden Formel für $\sigma_{y'}$:

$$\sigma_{y'} = \sqrt{\sigma_{\bar{y}}^2 + \sigma_b^2 (x_i - \bar{x})^2} \,. \qquad (5.10)$$

Die Größe $\sigma_{y'}$ wird üblicherweise als "Standardfehler der y-Koordinate eines Punktes auf der Regressionsgeraden" oder kurz als "Standardfehler der Regressionsgeraden" bezeichnet. $\sigma_{\bar{y}}$ ist der "Standardfehler der y-Koordinate des Schwerpunkts" und stellt eine Hilfsgröße zur Berechnung von $\sigma_{y'}$ dar. Mit σ_b bezeichnen wir den Standardfehler des Regressionskoeffizienten, mit x_i einen speziellen Beobachtungswert der unabhängigen Variablen und mit $\bar{x}$ das arithmetische Mittel aus diesen Werten. Gehen wir wieder davon aus, daß der Wert der Grundgesamtheit $\sigma_{y'}$ unbekannt ist und wir für ihn einen unverzerrten Schätzwert aus der Stichprobe ermitteln müssen, so entsteht aus der Weiterentwicklung der Gleichung (5.10) der folgende Ausdruck:

$$\hat{\sigma}_{y'} = \sqrt{\hat{\sigma}_{\bar{y}}^2 + \hat{\sigma}_b^2 (x_i - \bar{x})^2} \,. \qquad (5.11)$$

Formt man diese Beziehung weiter um, so daß in dem Ausdruck für $\hat{\sigma}_{y'}$ nur noch Größen erscheinen, die schon für die Berechnung des Standardfehlers des Regressionskoeffizienten erforderlich waren, dann erhält man[1)]

$$\hat{\sigma}_{y'} = \frac{\hat{\sigma}_{y.x}}{\sqrt{n}} \sqrt{1 + \frac{(x_i - \bar{x})^2}{\hat{\sigma}_x^2}} \qquad (5.12)$$

1) Zur Darstellung der Formeln vgl. auch E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 229 u. M. EZEKIEL u. K.A. FOX, a.a.O., S. 287, wo allerdings andere Symbole verwendet werden.

$$= \frac{s_{y \cdot x}}{\sqrt{n-2}} \sqrt{1 + \frac{(x_i - \bar{x})^2 (n-1)}{s_x^2 \cdot n}} \quad . \qquad (5.13)$$

Der wesentliche Unterschied zwischen $\hat{\sigma}_{y \cdot x}$ und $\hat{\sigma}_{y'}$ liegt darin begründet, daß $\hat{\sigma}_{y \cdot x}$ unabhängig von den Werten der unabhängigen Variablen für die gesamte Regressionsfunktion dieselbe Höhe aufweist, während $\hat{\sigma}_{y'}$ je nach Höhe der x-Werte unterschiedliche Werte annimmt. Dieser Tatsache wird durch die Aufnahme von x_i in die Formel für $\hat{\sigma}_{y'}$ Rechnung getragen. Die mit Hilfe von $\hat{\sigma}_{y'}$ ermittelten Vertrauensgrenzen verlaufen dadurch nicht parallel zur Regressionsgeraden, sondern bilden Kurven, wie in Abbildung 5.4 dargestellt wird. Diese Kurven nähern sich mit zunehmenden x_i-Werten der Regressionsgeraden, erreichen beim Erwartungswert $\bar{x}$ der x_i ihren geringsten Abstand von der Regressionslinie und streben von diesem Punkt aus wieder weiter auseinander.[1)]

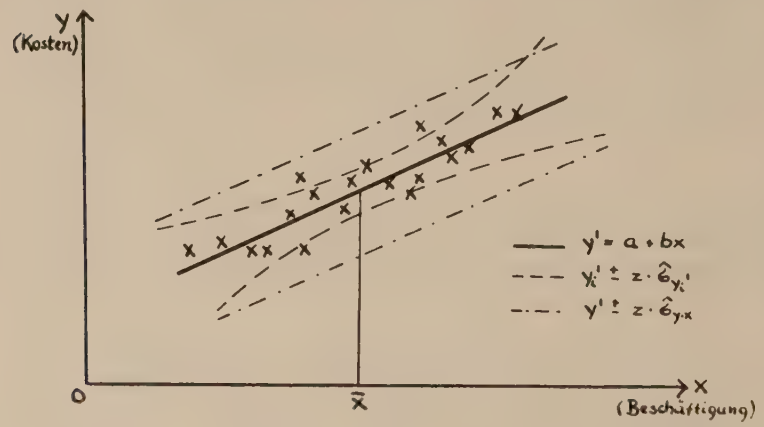


Abb. 5.4: Vertrauensgrenzen einer Regressionsgeraden, gebildet anhand des Standardfehlers der Schätzung $\hat{\sigma}_{y \cdot x}$ und des Standardfehlers der Regressionsgeraden $\hat{\sigma}_{y'}$.

1) Vgl. auch den Klammerausdruck $(x_i - \bar{x})^2$, der in allen Formeln für den Standardfehler der Regressionsgeraden erscheint.

Wir wollen hier auf die Durchrechnung eines Zahlenbeispiels verzichten, da lediglich gezeigt werden sollte, wie der Verlauf der auf $\hat{\sigma}_y$ basierenden Vertrauensgrenzen im Verhältnis zu den auf $\hat{\sigma}_{y.x}$ basierenden aussieht.[1] Anhand der Abbildung 5.4 oder der Berechnung von Konfidenzintervallen lassen sich im konkreten Fall dann Aussagen folgender Art treffen:
Die d u r c h s c h n i t t l i c h e n Kosten oder Verbrauchsmengen, d.h. die auf Grund der Regressionsfunktion geschätzten Werte, die einem bestimmten x-Wert entsprechen, liegen mit einer bestimmten, von z angegebenen Wahrscheinlichkeit in den Grenzen $y'_i \pm z \cdot \hat{\sigma}_{y'_i}$. Hierdurch lassen sich einmal zukünftige Verbrauchsmengen oder Kosten auf Grund von geplanten Werten ihrer Einflußgrößen schätzen, zum anderen aber auch Kostenkontrollen über die Abweichungen der Regreßwerte von den Beobachtungswerten durchführen.[2]

Schließlich soll noch kurz auf den "Standardfehler einer individuellen Prognose" (standard error of an individual forecast) eingegangen werden.[3] Dieses Streuungsmaß ist zur Bestimmung eines Konfidenzintervalls um einen bestimmten Regreßwert zu verwenden, wenn die Fragestellung etwas anders lautet als bisher. Der Standardfehler der Regressionsgeraden $\sigma_{y'_i}$ wurde zur Berechnung der Vertrauensgrenzen um die Regressionsgerade herangezogen, wenn man wissen wollte, in welchem Bereich die d u r c h s c h n i t t -

---

1) Graphische Darstellungen, die auf Grund konkreter Zahlenbeispiele erstellt wurden, findet der an weiteren Einzelheiten interessierte Leser bei E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 230.; M. EZEKIEL u. K.A. FOX, a.a.O., S. 139 u. S. 289.; F.E. CROXTON u. D.J. COWDEN, a.a.O., S. 492 u. S. 505 ff.

2) Vgl. den Abschnitt 6.2.

3) Vgl. M. EZEKIEL u. K.A. FOX, a.a.O., S. 318 ff.

l i c h e n Kosten oder Verbrauchsmengen bei einem bestimmten Wert x_i liegen. Dadurch sollte auf Grund der Stichprobenschätzwerte y'_i ein Konfidenzintervall für den wahren Regreßwert Y'_i der Grundgesamtheit abgesteckt werden. Stellt man jedoch die Frage, innerhalb welcher Grenzen man einen i n d i v i d u e l l e n Kosten- oder Verbrauchswert Y_i der Grundgesamtheit vermuten darf, dann genügt es nicht mehr, diesen Vertrauensbereich allein auf der Basis von $\sigma_{y'_i}$ zu berechnen, da zu diesem Schätzfehler ein weiterer hinzutritt, wenn - was vorausgesetzt wurde - der wahre Regreßwert Y'_i der Grundgesamtheit unbekannt ist. In diesem Fall setzt sich die Gesamtstreuung für Y_i aus der Streuung um das Y'_i der Grundgesamtheit - $\sigma_{y'_i}$ = Streuung der Einzelwerte um die Regressionsgerade der Grundgesamtheit - und aus der Streuung der Regressionsgeraden selbst - $\sigma_{y.x}$ - zusammen.[1] Aus der Addition beider Streuungsmaße ergibt sich der Standardfehler einer individuellen Prognose nach der Formel

$$\hat{\sigma}_F = \sqrt{\hat{\sigma}^2_{y.x} + \hat{\sigma}^2_{y'_i}} \quad . \qquad (5.14)$$

Aus der Gleichung (5.14) läßt sich unschwer erkennen, daß $\hat{\sigma}_F$ in aller Regel größer ist als $\hat{\sigma}_{y'_i}$. Die Frage, wie groß ein einzelner (spezieller), nicht auf der Regressionsgeraden liegender Kostenwert bei Annahme eines bestimmten x_i-Wertes ist, führt damit bei gleichem statistischem Sicherheitsgrad zu breiteren Konfidenzintervallen, oder bei gleichem Abstand der Vertrauensgrenzen zur Regressionslinie zu einer mit einer geringeren Wahrscheinlichkeit zu treffenden Aussage.

---

1) Vgl. M. EZEKIEL u. K.A. FOX, a.a.O., S. 319.

5.7 Regressionsgleichungen als Grundlagen für ein betriebliches Matrizenmodell

In den vorhergehenden Abschnitten wurde bereits mehrfach auf die Bedeutung der Bereitstellung von Unterlagen für kurzfristige Entscheidungen im Unternehmensbereich hingewiesen.[1] Es wurde auch erwähnt, daß die herkömmlichen Verfahren der Kostenrechnung in der Regel nicht in der Lage sind, geeignetes Zahlenmaterial hierfür zur Verfügung zu stellen, da sie zu wenig Kenntnisse über Abhängigkeiten oder Gesetzmäßigkeiten zwischen Kosten oder Verbrauchsmengen und den ihre Höhe beeinflussenden Faktoren vermitteln.

Im Verlauf dieser Arbeit wurde gezeigt, auf welche Weise die Regressions- und Korrelationsanalyse dazu dienen kann, diese Abhängigkeiten in mathematischer Form auszudrücken. In der Eisen- und Stahlindustrie hat man sich diese Formeln zunutze gemacht und versucht, auf ihnen aufbauend eine Art "m a t h e m a t i s c h e s B e t r i e b s m o d e l l" zu entwickeln. Da solche Modelle in der Zukunft große praktische Bedeutung erlangen könnten, sollen hier in aller Kürze einige Gedanken zu dieser neuen Form eines Rechnungswesens wiedergegeben werden.

Für kurzfristige Entscheidungen im Produktionsbereich sind insbesondere Kenntnisse über die vielfältigen innerbetrieblichen Verflechtungen und technischen Gegebenheiten erforderlich. Kann man ihre Struktur aufdecken und in mathematischen Formeln erfassen, so bietet sich die Möglichkeit, diese Formeln von Fall zu Fall auf das augenblickliche oder geplante Betriebsgeschehen zu beziehen - durch den Einsatz

1) Vgl. S. 97 f. und S. 118 ff.

der in Frage kommenden Zahlen - und eine Art von "Schubladenplanung" aufzubauen, die je nach der betrieblichen Situation Entscheidungsalternativen liefert.[1] Die mathematische Formulierung eines solchen "Betriebsmodells" wird zweckmäßigerweise in Form eines Matrizenmodells vorgenommen. Will man zum Beispiel für die Kosten eines Hochofens einen mathematischen Ausdruck ermitteln, dann bietet sich in der Darstellungsform der Matrizenrechnung folgende Formulierung an:[2]

$$K = \mathbf{p}' \cdot \mathbf{H} \cdot \mathbf{x} \qquad (5.15)$$

oder $K = \mathbf{p}' \cdot \mathbf{y}$, wobei gilt: $\mathbf{y} = \mathbf{H} \cdot \mathbf{x}$. (5.16)

In dieser Gleichung haben die Symbole folgende Bedeutung:

K = gesamte Selbstkosten,

$\mathbf{p}'$ = Vektor für die Preise der Verbrauchsmengen, wobei gilt:

$$\mathbf{p}' = (p_1, p_2, \ldots, p_n),$$

$\mathbf{x}$ = Vektor für das im betrachteten Zeitraum gefertigte Programm (Mengenvektor), wobei gilt:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

---

1) Vgl. R. WARTMANN, H.-J. KOPINECK u. W. HANISCH: Funktionales Arbeiten in Kostenrechnung und Planung mit Hilfe von Matrizen. "Archiv für das Eisenhüttenwesen". 31 (1960) 7, S. 441 ff., hier S. 441.

2) Ebenda. S. 446.

$\mathbf{H}$ = Matrix für das technische Betriebsgeschehen, d.h. die Arbeitsweise der betrachteten Abteilung (Kostenstrukturmatrix), wobei gilt:

$$\mathbf{H} = \begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & h_{22} & \cdots & h_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ h_{m1} & h_{m2} & \cdots & h_{mn} \end{pmatrix}$$

$\mathbf{y}$ = Vektor, der die Verbrauchsmengen angibt, wobei gilt:

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{pmatrix}$$

"Die Matrizengleichung sagt nun aus: Die Aufwendungen sind eine Folge des Arbeitsprogrammes $\mathbf{x}$, das mit Hilfe der Gegebenheiten $\mathbf{H}$ des Hochofenwerkes durchgeführt wird. Hierbei steht die Gleichung $\mathbf{y} = \mathbf{H} \cdot \mathbf{x}$ stellvertretend für die .. einzelnen Gleichungen, die im Grundsätzlichen ja ge-

nau so gebaut sind, nämlich

abhängige Variable y = Faktor • Einflußgröße x."[1]

Durch Multiplikation des Vektors **Y** mit dem Preisvektor **P'** ergeben sich dann die K o s t e n.

Dieser grobe Überblick über die Grundlage des Modells mag genügen, um den Rahmen für die folgenden Aussagen abzustecken.[2] Die Matrix **H** für das Betriebsgeschehen repräsentiert die Abhängigkeiten der Mengenkomponenten jeder einzelnen Kostenart von bestimmten Einflußfaktoren (K o s t e n s t r u k t u r m a t r i x). Diese Abhängigkeiten in mathematischer Form zu liefern ist die Aufgabe, die im Rahmen dieses Modells an die Regressionsanalyse gestellt wird. Die von ihr ermittelten Beziehungen bilden damit gewissermaßen "das Grundgerüst dieser Planungsrechnung".[3]

In den zitierten Veröffentlichungen findet man zahlreiche Beispiele für die Ermittlung von Abhängigkeiten zwischen Verbrauchsmengen und ihren Einflußfaktoren mit Hilfe der einfachen, aber auch der multiplen Regressionsanalyse. Wir können an dieser Stelle auf die Wiedergabe dieser Beziehungen oder auf die Berechnung eigener Zahlenbeispiele verzichten, da die Bestimmung der Regressionsparameter eben nach den Methoden erfolgt, die in dieser Arbeit schon in aller Ausführlichkeit beschrieben worden sind. Dem Verfasser kam es lediglich darauf an zu zeigen, daß die Ergebnisse einer Regressionsanalyse keineswegs nur für einen be-

---

1) R. WARTMANN, H.-J. KOPINECK u. W. HANISCH, a.a.O., S. 444.

2) Weitere Einzelheiten findet der interessierte Leser bei: R. WARTMANN, H.-J. KOPINECK u. W. HANISCH, a.a.O., S. 444 ff.; M. STEFFEN u. V. STEINECKE: Einflußgrößenrechnung zur Kostenplanung eines kontinuierlichen Feinstahlwalzwerkes mit Matrizen. "Stahl und Eisen". 82 (1962) 3, S. 155 ff.; G. LASSMANN, a.a.O., S. 72 ff.

3) R. WARTMANN: Steuern, Lenken, Planen. Möglichkeiten in einem gemischten Hüttenwerk. "Hoesch-Berichte aus Forschung und Entwicklung unserer Werke". O.J. (1966), 2, S. 20 ff., hier S. 22.

stimmten isolierten Rechnungszweck von Nutzen sind, sondern daß gerade für den Aufbau eines umfassenden Rechnungssystems die durch die Regressionsfunktionen wiedergegebenen Kostenabhängigkeiten von großer Bedeutung sind.

5.8 Zusammenfassende Beurteilung der statistischen Kostenanalyse

In den Abschnitten dieses 5. Kapitels wurde anhand einer Reihe von Beispielen gezeigt, wie vielfältig die Anwendungsmöglichkeiten für die Regressionsrechnung auf dem Gebiet der Kostenanalyse sind und welchen Nutzen der Kostenrechner insbesondere aus der Durchführung von multiplen, schrittweisen Regressionsanalysen ziehen kann. Wir haben gesehen, wie mit diesem Verfahren häufig eine Berücksichtigung zukünftiger betrieblicher Situationen gelingt und wie dadurch die Flexibilität regressionsanalytischer Kostenplanungen erhöht werden kann. Durch die Einbeziehung von Elementen der Wahrscheinlichkeitsrechnung bietet sich über die Berechnung von Konfidenzintervallen und die Durchführung von Tests die Möglichkeit, die ermittelten Durchschnittswerte "statistisch zu sichern", was besonders für die Zuverlässigkeit der Schätzung zukünftiger Daten auf Grund der Regreßwerte von Bedeutung ist. Schließlich wurde darauf hingewiesen, wie mit Hilfe der Regressionsgleichungen, die Abhängigkeiten oder Gesetzmäßigkeiten zwischen Kosten und ihren Einflußgrößen wiedergeben, der Versuch zur Aufstellung eines mathematischen Betriebsmodells unternommen werden kann, ein Vorhaben, das zweifellos zu einer stärkeren Objektivierung in der Kostenrechnung beitragen wird.

Daneben wurde aber auch gezeigt, welche Probleme die Anwendung der Regressionsanalyse bei Kostenuntersuchungen

mit sich bringt. Jedoch sollte man bei allen betriebswirtschaftlich-theoretischen Erörterungen dieser Problematik bei der praktischen Ermittlung von Schätzwerten für Gesamtkosten, Grenzkosten, Verbrauchsmengen und ähnlichen Größen nicht übersehen, daß die statistischen Verfahren - setzt man den Einsatz von programmierbaren Tischrechnern oder elektronischen Datenverarbeitungsanlagen voraus - relativ einfach und schnell und damit kostengünstig durchzuführen sind. Insoweit kann der Verfasser D. SCHNEIDERs Ansicht nicht zustimmen, daß "statistische Kostenuntersuchungen ... zeitraubend und teuer [sind] ".[1] Sicherlich kann in bestimmten Fällen durch erforderliche Aufbereitungsarbeiten ein so großer zusätzlicher Arbeitsaufwand entstehen, daß die rechentechnischen Vorteile, die die Regressionsanalyse durch den Einsatz programmierter Rechenanlagen besitzt, kompensiert werden. Häufig wird man aber davon ausgehen können, daß dieses statistische Verfahren - auch ohne ins Detail gehende, außerhalb der Regressionsanalyse durchgeführte gesonderte Kostenuntersuchungen - praktisch annehmbare Resultate liefert,[2] wobei möglicherweise in der Praxis diese mit geringerem Aufwand ermittelten Werte den auf Grund umfangreicherer Untersuchungen zu gewinnenden exakteren Resultaten vorgezogen werden.

---

1) D. SCHNEIDER, Kostentheorie ..., a.a.O., S. 682.

2) So auch A. MATZ, a.a.O., S. 87.

6. Kapitel

KOSTENKONTROLLE

6.1 Einführung

In den bisherigen Kapiteln dieser Arbeit wurde schon wiederholt auf die Bedeutung der Kostenkontrolle - insbesondere der Gemeinkosten - im modernen Rechnungswesen hingewiesen. Dabei wurde auch kurz - so im Rahmen der Gemeinkostenplanung und bei der Aufstellung von Kostenbudgets - auf die Möglichkeiten zur Kostenkontrolle eingegangen, die sich durch die Anwendung der Regressions- und Korrelationsanalyse ergeben.

Grundsätzlich können und sollten alle mit Hilfe der Regressionsanalyse geplanten oder budgetierten Kosten auch während oder nach Ablauf des Planungszeitraumes überprüft werden, denn jede Planung hat nur dann ihren Sinn erfüllt, wenn später das Geplante mit dem tatsächlich eingetretenen Ist verglichen, d. h, kontrolliert wird.

Im ersten Abschnitt dieses Kapitels soll deshalb am Beispiel eines innerbetrieblichen Zeitvergleichs der Kosten gezeigt werden, inwieweit sich die mit Hilfe der Regressionsanalyse ermittelten Ergebnisse für Kontrollzwecke eignen. In diesem Zusammenhang ist es interessant festzustellen, daß die Kostenrechnung und die statistischen Analysen manche gemeinsame Elemente besitzen. Wenn in der Statistik auf Grund einer Gesamtheit oder einer Stichprobe Mittelwerte und Streuungsmaße ermittelt werden, so geht die Kostenrechnung ganz ähnlich vor, indem auch sie Gesamtheiten oder Teile von Gesamtheiten bezüglich sich wiederholender kostenverursachender Vorgänge untersucht, "Plankosten oder Kostenvorgaben (die die Aufgabe eines

Durchschnitts- oder Mittelwertes erfüllen) errechnet und schließlich Abweichungen als Differenz zwischen Soll (= Mittelwert) und Ist errechnet".[1] Bei den Untersuchungen zur Kostenkontrolle wird sich herausstellen, daß die auf die Regressionsergebnisse anwendbaren statistischen Methoden der Bildung von Konfidenzintervallen über die Aufstellung von "Kostenkontrollkarten" zu einem Grundmodell des "management by exception" führen können.

Der zweite Abschnitt,in dem auf den zwischenbetrieblichen Kostenvergleich eingegangen wird, soll vor allem auf Probleme im Zusammenhang mit der Wahl der geeigneten Funktionsform aufmerksam machen. Da bisher nur lineare Regressionsfunktionen gewählt wurden, soll im Abschnitt 6.3 auf ein Rechenbeispiel eingegangen werden, an dem sich die Problematik der Wahl der richtigen Kurvenform für die Regressionsfunktion demonstrieren läßt.

6.2 Innerbetrieblicher Zeitvergleich

6.2.1 Bildung eines Konfidenzintervalls um die Regressionsgerade

Beim innerbetrieblichen Zeitvergleich werden die einzelnen Kostenarten in ihrer geplanten und ihrer tatsächli-

1) H.-M. SCHÖNFELD: Die Anwendung ..., a.a.O., (1968), S. 245. Vgl. in diesem Zusammenhang auch Z.S. ZANNETOS: Standard Costs as a First Step to Probabilistic Control: A Theoretical Justification, an Extension and Implications. "The Accounting Review". 39 (1964) 2, S. 296ff., der auch auf diese Ähnlichkeit zwischen Kostenrechnung und statistischer Analyse hingewiesen hat.

chen Höhe vergleichend nebeneinander gestellt,[1] wie das zum Beispiel in der flexiblen Plankostenrechnung durch die Gegenüberstellung von Sollkosten und Istkosten geschieht. Am Ende der Rechnungsperiode (Woche, Monat, Jahr) will man feststellen, ob die Kosten in einer bestimmten Kostenstelle "im Rahmen des Üblichen liegen" oder ob sie "außer Kontrolle" sind, sei es im günstigen oder ungünstigen Sinn. Dabei wird der "Rahmen des Üblichen" durch die mit Hilfe der Regressionsanalyse ermittelten Schätzwerte für die Kosten abgesteckt.

Zur Verdeutlichung der Vorgehensweise beziehen wir uns auf ein Rechenbeispiel einer Regression mit einer einzigen unabhängigen Variablen. In diesem Fall sind die Ergebnisse graphisch leichter zu interpretieren, da die Relation zwischen abhängiger und unabhängiger Variabler im zweidimensionalen Koordinatensystem darstellbar ist. Jedoch gelten die im folgenden getroffenen Aussagen sinngemäß auch für multiple Regressionen.

In den der Analyse folgenden Beobachtungszeiträumen werden nun die tatsächlich entstandenen Kosten bzw. Verbrauchsmengen - je nachdem, ob lediglich mit der Mengenkomponente oder unter Einbeziehung des Preises mit Kosten gerechnet wird - mit den auf Grund der Regreßwerte ermittelten Kosten verglichen. Um diesen Vergleich zu veranschaulichen, legen wir das in Tabelle 6.1 wiedergegebene Zahlenmaterial zugrunde.[2]

---

1) Vgl. E. SCHMALENBACH, a.a.O., S. 436.

2) Die Zahlenwerte wurden entnommen aus E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 43 Tabelle 1. Im Original wird eine Relation zwischen Produktionsvolumen und Grundfonds untersucht, wir wollen für unsere Zwecke die y_i-Werte als "Kosten" und die x_i-Werte als "Beschäftigung" interpretieren.

Periode	y_i	x_i	Periode	y_i	x_i
1	208	12	31	253	32
2	214	16	32	236	33
3	219	16	33	247	33
4	222	18	34	253	33
5	222	21	35	258	33
6	227	23	36	245	34
7	232	23	37	251	34
8	239	23	38	258	34
9	231	24	39	251	36
10	232	25	40	255	36
11	242	25	41	259	36
12	236	26	42	265	36
13	242	26	43	253	37
14	247	26	44	259	37
15	236	28	45	272	38
16	240	28	46	265	39
17	242	28	47	275	41
18	247	28	48	272	44
19	252	28	49	279	44
20	240	29	50	283	47
21	244	29	51	287	48
22	240	30	52	296	54
23	242	30			
24	245	30			
25	249	30			
26	255	30			
27	245	31			
28	247	31			
29	251	31			
30	245	32			

Tabelle 6.1: Zahlenwerte für Kosten (y_i) und Beschäftigung (x_i) für 52 Beobachtungsperioden, geordnet nach der Höhe der Beschäftigung.

Auf Grund dieses Datenmaterials haben wir die im Schaubild 6.1 eingezeichnete Regressionsgerade

$$y' = 183{,}06 + 2{,}095\ x$$

als Relation zwischen den Kosten (y) und ihrer Einflußgröße Beschäftigung (x) gefunden. Die in Abb. 6.1 durch ein "x" gekennzeichneten Punkte symbolisieren die b e o b a c h t e t e n Kosten y_i, während $y'_i = 245{,}9$ den bei einer Beschäftigung von $x_i = 30$ auf Grund der Regressionsfunktion ermittelten S c h ä t z w e r t für die Kosten repräsentiert. Bei diesem Beschäftigungsgrad haben wir fünf unterschiedliche Beobachtungswerte für y_i erhalten, nämlich:

$$y_{22} = 240,$$
$$y_{23} = 242,$$
$$y_{24} = 245,$$
$$y_{25} = 249,$$
$$y_{26} = 255.$$

Wir wollen annehmen, daß darüber hinaus während der Folgemonate bei einer Beschäftigung von $x_i = 30$ für y die Werte

$$y_{53} = 260 \text{ und } y_{54} = 258$$

beobachtet wurden. Daraufhin stellt sich die Frage, ob diese Werte noch im Rahmen des Üblichen liegen und durch irgendwelche dem Stichprobenmodell anzulastende Schwankungen erklärt werden können, oder ob teilweise so starke Abweichungen von der Norm (dem Durchschnitt) gegeben sind, daß eine weitergehende Untersuchung der Kosten zweckmäßig erscheint, weil man vermutet, daß sie durch betriebliche Maßnahmen (Verfahrensänderungen, Unwirtschaftlichkeiten

usw.) beeinflußt worden sind.[1]

Dazu gehen wir von der Voraussetzung aus, daß die Residuen $e_i = (y_i - y_i')$ voneinander unabhängig sind und eine Normalverteilung um die Regressionsgerade bilden, mit dem Mittelwert y_i' und der konstanten Standardabweichung $s_{y.x}$.[2] Anders formuliert lauten die Voraussetzungen unseres Modells: Homoskedasie, keine Autokorrelation und Normalverteilung.[3]

In diesem Fall können wir in Abb. 6.1 die Parallelen zur Regressionsgeraden einzeichnen[4] sowie die Dichtefunktion der standardisierten Normalverteilung[5] über dem für $x_i = 30$ ermittelten Schätzwert für die Kosten ($y_i' = 245{,}9$) errichten und können anhand dieser Kurve und den Tabellenwerten der Normalverteilung Aussagen über die Zuverlässigkeit der Schätzwerte treffen, wie im folgenden dargestellt wird.

Obwohl sich die Normalverteilung strenggenommen nur anwenden läßt, wenn eine binormalverteile Grundgesamtheit vor-

---

1) Vgl. auch E.E. COMISKEY: Cost Control by Regression Analysis. "The Accounting Review". 41 (1966) 2, S. 235 ff., hier S. 237.

2) Vgl. auch E. MANSFIELD u. H.H. WEIN: A Regression Control Chart for Costs. "Applied Statistics. A Journal of the Royal Statistical Society". 7 (1958) 1, S. 48 ff., hier S. 50. Zur Problematik der "Heteroskedasie" und der "Autokorrelation der Residuen" vgl. auch die Abschnitte 7.1 und 7.2.

3) $s_{y.x}$ wird üblicherweise als Standardabweichung der Residuen oder als Standardfehler der Schätzung bezeichnet. Zur ausführlichen Darstellung vgl. Abschnitt 5.6.4 und die Formel 5.9 auf S. 156.

4) Es wurde der Abstand von e i n e r Standardabweichung gewählt.

5) Die Größe des Standardfehlers der Residuen bestimmt, wie steil diese Kurve verläuft.

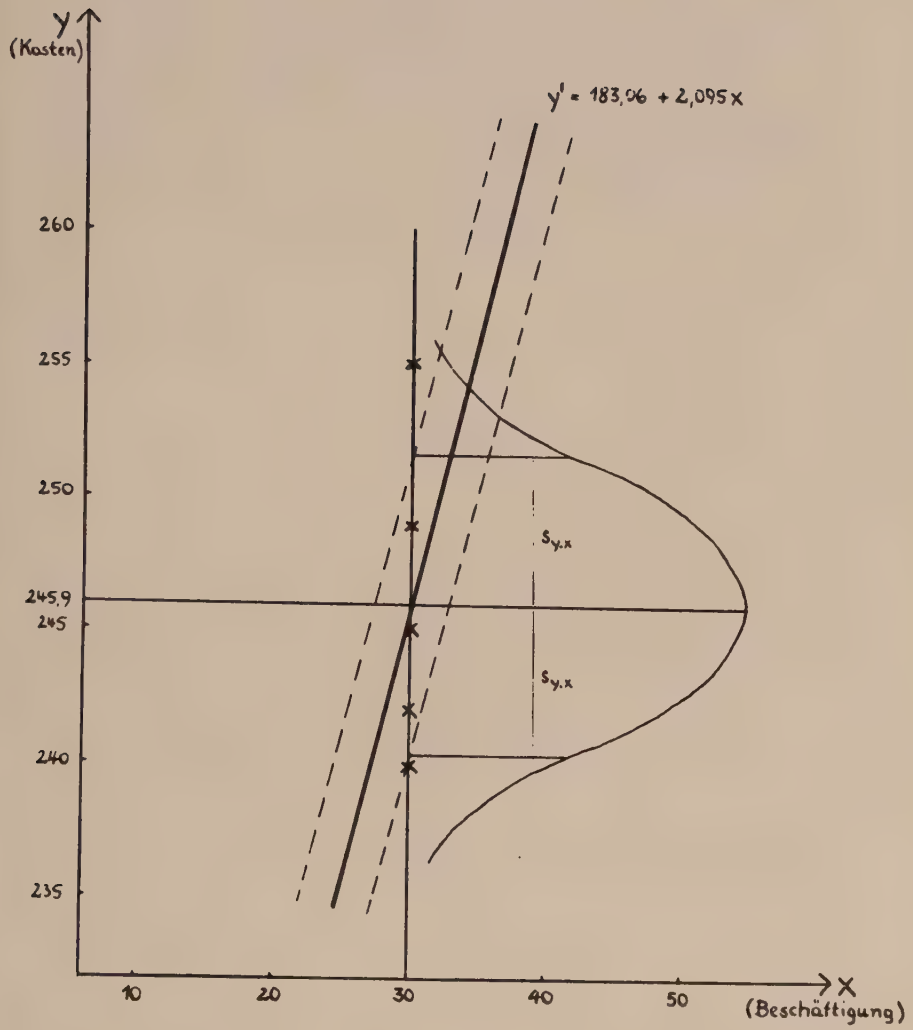


Abb. 6.1: Regressionsgerade einschließlich Konfidenzintervall im Abstand einer Standardabweichung und Glockenkurve der Normalverteilung.

liegt, d. h. wenn die Werte der unabhängigen Variablen x und die Werte der abhängigen Variablen y in der Grundgesamtheit normalverteilt sind und die Residuen um die Regressionsfunktion der Grundgesamtheit mit konstanter Standardabweichung $\sigma_{y.x}$ normalverteilt sind, läßt sich dennoch bei Abweichungen von diesen Voraussetzungen auf Grund des "Zentralen Grenzwertsatzes"[1] bei genügend großem Stichprobenumfang für die Verteilung der Residuen in der Stichprobe eine Normalverteilung zugrundelegen.

6.2.2 Aufstellung einer Kostenkontrollkarte

Man kann nun folgende Überlegungen anstellen: Da die Regreßwerte Mittelwerte (Durchschnittswerte) sind, die mit einer statistischen Wahrscheinlichkeit innerhalb eines bestimmten Intervalls streuen können,[2] läßt sich auf Grund einer festgestellten Diskrepanz zwischen einem beobachteten Wert y und dem mit Hilfe der Regressionsanalyse geschätzten Kosten- oder Verbrauchswert y' nicht ohne weiteres darauf schließen, daß die festgestellten (beobachteten) Werte "aus dem Rahmen des Üblichen" fallen und daß aus diesem Grund weitere Untersuchungen erforderlich sind, um die Ursachen dieser Abweichung festzustellen. Die Anwendung der Regressionsanalyse verlangt also "ein Denken in Toleranzen".[3] Diese auf den ersten Blick nachteilig wirkende Erscheinung erweist sich jedoch als Vorteil dieses Verfahrens, da diese Toleranzen b e r e c h n e t und somit bei Entscheidungen berücksichtigt werden können.

---

1) Vgl. z.B. A. HALD, a.a.O., S. 189.; H.-M. SCHÖNFELD: Die Anwendung ..., a.a.O., (1969), S. 20.; vgl. auch S. 147 f. dieser Arbeit.

2) Vgl. auch E. FÖRSTER: Zur Anwendung ..., a.a.O., S. 562.; E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 103.

3) Ebenda. S. 106.; E. FÖRSTER: Zur Anwendung ..., a.a.O., S. 563.

Darüber hinaus ist der Tatsache Rechnung zu tragen, daß die ermittelte Regressionslinie selbst eine S t i c h p r o b e n r e g r e s s i o n s l i n i e ist, die auf Grund dieses Stichprobencharakters gewissen zufälligen Schwankungen unterliegt.[1] Die Regressionsparameter a und b sind also Stichprobenschätzwerte für die "wahren" Regressionsparameter α und β der Grundgesamtheit und streuen in bestimmten Grenzen, wie im Abschnitt 5.6 dargestellt wurde.[2] Da die Parameter α und β und auch die Standardabweichung der Residuen $\sigma_{y.x}$ der Grundgesamtheit unbekannt sind, müssen wir mit Stichprobenschätzwerten für diese Größen rechnen. Wir wollen im folgenden die möglichen Schwankungen von a und b außer acht lassen und sie behandeln, als wenn es die wahren Werte der Parameter wären, d. h. wir nehmen bei der folgenden Untersuchung an, daß die Beziehungen $\hat{\alpha} = a$ und $\hat{\beta} = b$ korrekt sind.

Wir haben damit als erstes Ergebnis festgestellt, daß erst bei "signifikanten" Abweichungen zwischen beobachteten und geschätzten Kosten gründliche Neuuntersuchungen der Kosten oder Verbrauchsmengen sinnvoll sind, da auf Grund der Voraussetzungen unseres Stichprobenmodells nur in diesem Fall mit betrieblich bedingten Ursachen für diese Abweichungen zu rechnen ist.

Was bedeutet aber der Begriff "signifikante Abweichungen"? Zur Verdeutlichung soll das folgende Schaubild dienen, das aus der Abb. 6.1 entwickelt wurde.[3]

---

1) Vgl. E. MANSFIELD u. H.H. WEIN, a.a.O., S. 50 f.

2) Vgl. S. 155 ff.

3) Mit H_0 wird die Nullhypothese: keine Änderung des bestehenden Zustandes, bezeichnet. Vgl. S. 69 ff., insbesondere aber S. 180 f., wo ausführlich auf die Definition von H_0 eingegangen wurde.

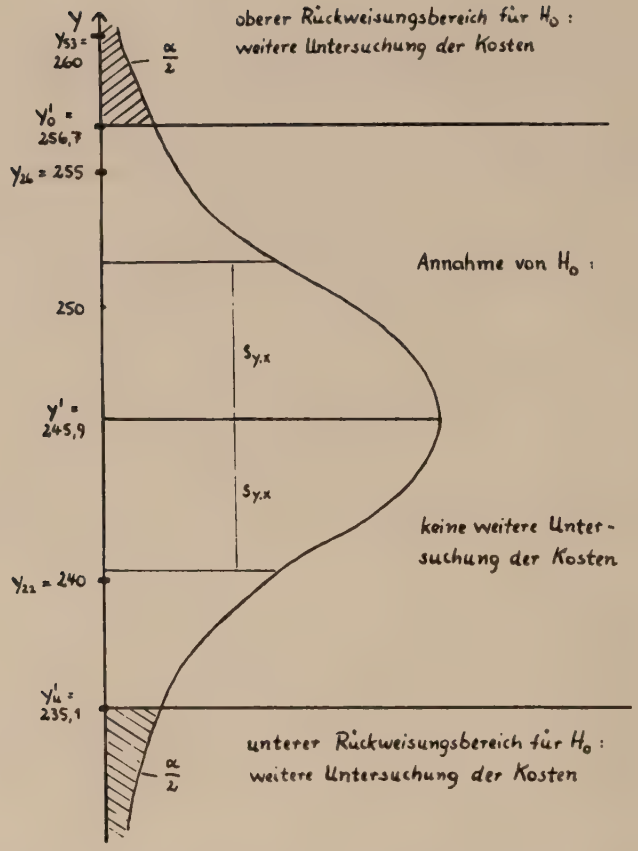


Abb. 6.2: Darstellung einer Kostenkontrollkarte anhand der Normalverteilung.

Wie auch in Abschnitt 5.6 dargestellt wurde, läßt sich durch Standardisierung der zweiparametrischen Normalverteilung jede normalverteilte Zufallsvariable X in eine Zufallsvariable Z überführen, die den Mittelwert

(Erwartungswert) $\mu = 0$ und die Varianz $\sigma^2 = 1$ besitzt.[1] Durch diese Transformation läßt sich eine Tabelle aufstellen - Tabelle der Standardnormalverteilung - in der sich für jede Normalverteilung mit gegebenem Mittelwert und gegebener Standardabweichung die gesuchte Wahrscheinlichkeit ablesen läßt.[2]
Dabei gilt für Z die Formel

$$Z = \frac{X - \mu}{\sigma} , \qquad (6.1)$$

oder auf unser Beispiel übertragen,[3]

$$z = \frac{y - y'}{s_{y.x}} , \qquad (6.2)$$

d. h. die Abweichung zwischen y und y' (Residuum) wird als Vielfaches der Standardabweichung der Residuen (Standardfehler der Schätzung) gemessen, um den tabellierten z-Wert zu ermitteln.

Setzen wir die Zahlenwerte unseres Beispiels für die beobachteten Kosten y_{22} bis y_{26} und y_{53} bis y_{54} sowie den auf Grund des Datenmaterials der Tabelle 6.1 errechneten Wert für den Standardfehler der Schätzung $s_{y.x}$ in die Gleichung (6.2) ein,[4] so erhalten wir folgende z-Werte:

---

1) Vgl. z. B. auch J. PFANZAGL: Allgemeine Methodenlehre der Statistik II. Höhere Methoden unter besonderer Berücksichtigung der Anwendungen in Naturwissenschaften, Medizin und Technik. 2., verb. Aufl. Berlin 1966. S. 27 ff.

2) Vgl. die Tabellen 1* und 2* der Standardnormalverteilung im Anhang II, S. 284 f.

3) Die Kosten einer bestimmten Beobachtung (y_i) werden als normalverteilte Zufallsvariable angesehen. Vgl. auch S. 146 f.

4) Der Standardfehler der Schätzung errechnet sich nach der in Gleichung 5.9 angegebenen Formel. Zur vereinfachten Berechnung eines unverzerrten Schätzwertes $\hat{\sigma}_{y.x}$ vgl. z. B. E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 199.; M. EZEKIEL u. K.A. FOX, a.a.O., S. 123 f.

$$y_{22}: \quad z_{22} = \frac{240 - 245{,}9}{5{,}5} = -1{,}07,^{1)}$$

$$y_{23}: \quad z_{23} = \frac{242 - 245{,}9}{5{,}5} = -0{,}71,$$

$$y_{24}: \quad z_{24} = \frac{245 - 245{,}9}{5{,}5} = -0{,}16,$$

$$y_{25}: \quad z_{25} = \frac{249 - 245{,}9}{5{,}5} = 0{,}56,$$

$$y_{26}: \quad z_{26} = \frac{255 - 245{,}9}{5{,}5} = 1{,}65,$$

$$y_{53}: \quad z_{53} = \frac{260 - 245{,}9}{5{,}5} = 2{,}56,$$

$$y_{54}: \quad z_{54} = \frac{258 - 245{,}9}{5{,}5} = 2{,}20.$$

Wir nehmen an, daß die Aussagen über die Zuverlässigkeit der Regreßwerte mit einem statistischen Konfidenzniveau von $1 - \alpha = 95\%$ getroffen werden sollen, d. h. man will erst dann weitere Kosten- oder Verbrauchsuntersuchungen durchführen, wenn mit einer Wahrscheinlichkeit von 95% feststeht, daß beobachtete Abweichungen nicht auf Zufallsschwankungen, sondern auf tatsächlich betrieblich bedingte Ursachen zurückzuführen sind. Für das gewählte Signifikanzniveau $\alpha = 0{,}05$ liest man aus der Tabelle der Standardnormalverteilung[2)] für den hier durchgeführten zweiseitigen Test[3)] einen Wert für $z_{0,95} = 1{,}96$ ab. Beim Vergleich dieses z-Wertes mit den auf Grund der Stichprobenergebnisse errechneten z-Werten zeigt sich folgendes:

---

1) Das negative Vorzeichen sagt aus, daß der beobachtete Wert y_{22} rechts vom Mittelwert y' liegt. (Vgl. Abb. 6.2).

2) Die theoretisch exakte Studentverteilung kann bei dem relativ großen Stichprobenumfang von $n = 52$ durch die Normalverteilung ersetzt werden. Vgl. Tabelle 2* im Anhang II, S. 285.

3) Es werden sowohl Kostenabweichungen nach unten als auch nach oben - vom Mittelwert aus gesehen - untersucht.

Die z-Werte z_{22} bis z_{26} sind kleiner als der kritische z-Wert, der dem geforderten statistischen Sicherheitsgrad von 0,95 entspricht. Das bedeutet, daß sich die Abweichungen zwischen den Werten y_{22} bis y_{26} und dem Schätzwert $y' = 245,9$ noch durch zufällige Einflüsse (Streuung) erklären und nicht auf Änderungen im Beobachtungsdatenmaterial schließen lassen. Dagegen sind die Werte z_{53} und z_{54} größer als der kritische Wert $z_{0,95}$. Hier läßt sich die Diskrepanz zwischen den beobachteten Kosten und den geschätzten Kosten durch Zufallsschwankungen allein nicht mehr erklären, d. h. es liegt eine signifikante Abweichung zwischen beiden Größen vor. Mit einer Wahrscheinlichkeit von mindestens 95% läßt sich aussagen, daß in diesem Fall andere Ursachen (Verfahrensänderungen, Unwirtschaftlichkeiten und ähnliche, betrieblich bedingte Einflüsse) mitgewirkt haben müssen.

Anstelle des Vergleichs der z-Werte lassen sich durch die Bildung von Kontrollgrenzen - wie in Abb. 6.2 geschehen - auch die y-Werte direkt mit dem geschätzten y'-Wert vergleichen. Bei einem Konfidenzniveau von $1 - \alpha = 0,95$ und einer Standardabweichung von 5,5 müssen 95% aller beobachteten y-Werte - es gilt immer $x_i = 30$ - in den Bereich

$$y' \pm z \cdot s_{y.x} = 245,9 \pm 1,96 \cdot 5,5 = 245,9 \pm 10,8$$

fallen, so daß die untere Kontrollgrenze y'_u bei 235,1 und die obere Kontrollgrenze y'_o bei 256,7 liegt. Es gilt folgende Beziehung:

$$W(y' - 1,96 \cdot s_{y.x} < y \leq y' + 1,96 \cdot s_{y.x}) = 0,95,$$

$$W(235,1 < y \leq 256,7) = 0,95.$$

Auch auf diese Weise zeigt sich, wie man aus Abb. 6.2 direkt ablesen kann, daß die Abweichungen zwischen y' und

den Werten y_{22} bis y_{26} durch Zufallsschwankungen erklärt werden können, nicht jedoch die Abweichungen, die die beobachteten Werte y_{53} und y_{54} betreffen. Diese Werte fallen in den schraffierten Bereich und führen zu einer Ablehnung der Nullhypothese.[1)]

6.2.3 Testen der Nullhypothese

Bei den im vorigen Abschnitt durchgeführten Untersuchungen handelt es sich um eine statistische Hypothesenprüfung, wie sie bereits im Abschnitt 4.4.4 beim Test des Korrelationskoeffizienten durchgeführt wurde. Auch im Rahmen der Kostenkontrolle will man feststellen, ob ein Stichprobenergebnis in Übereinstimmung mit einer bestimmten Hypothese steht, wobei die Aussagen mit einer angebbaren statistischen Wahrscheinlichkeit getroffen werden können.

Zum Test der Nullhypothese in Verbindung mit einer Kostenkontrollkarte, die durch die Regressionsergebnisse gewonnen wurde, sind folgende Erläuterungen erforderlich. Obwohl der mit Hilfe der Regressionsanalyse ermittelte Schätzwert y' für die tatsächlichen Kosten selbst ein Wert ist, der auf Grund einer Stichprobe gefunden wurde,[2)] dient er hier als G r u n d l a g e f ü r d i e H y p o t h e s e , während das zu testende Stichprobenergebnis durch den für die tatsächlichen Kosten y beobachteten Wert repräsentiert wird. Anders formuliert, die Grundlage unserer Untersuchung bildet die Regressionsfunktion oder die auf ihrer Grundlage errechneten Schätzwerte y', das zu testende Stichprobenergebnis bildet der während des Produktionsprozesses beobachtete Kostenwert, zum Beispiel der Wert y_{53} oder y_{54}.

---

1) Vgl. den folgenden Abschnitt 6.2.3.

2) Vgl. auch den Abschnitt 5.6.; und H.-M. SCHÖNFELD: Die Anwendung ..., a.a.O., (1968), S. 248.

Als Nullhypothese H_0 hatten wir folgende These aufgestellt:[1] Es ist keine Änderung des bestehenden Zustandes eingetreten, d. h. die beobachteten Werte stammen aus der gleichen Grundgesamtheit wie y'. Auf das konkrete Problem übertragen bedeutet diese Aussage: Es sind keine Änderungen im Produktionsverfahren eingetreten oder Unwirtschaftlichkeiten oder andere betriebliche Einflüsse aufgetreten. Bei einem Konfidenzniveau von $1 - \alpha = 0{,}95$ beträgt die Wahrscheinlichkeit, diese Hypothese auf Grund der beobachteten Werte fälschlicherweise abzulehnen - und die Alternativhypothese anzunehmen[2] - 5%. Diese Wahrscheinlichkeit bezeichnet man üblicherweise als Risiko eines α-Fehlers oder Fehlers erster Art.

Die Höhe des z-Wertes, d. h. des geforderten statistischen Signifikanzniveaus und damit auch die Höhe des α-Fehlers, hängt weitgehend von dem "Sicherheitsbedürfnis" des für die Entscheidung über die Durchführung einer weitergehenden Untersuchung Verantwortlichen ab, unterliegt also in gewisser Weise subjektiven Momenten.[3] Allerdings wird man in der Regel mit z-Werten arbeiten, die einem Konfidenzniveau zwischen 95% und 99% entsprechen, so daß - gleichen Standardfehler der Schätzung vorausgesetzt - erhebliche Abweichungen in jedem Fall außerhalb der Kontrollgrenzen liegen werden und sich lediglich bei kleineren Abweichungen - je nach gefordertem Sicherheitsgrad - Entscheidungen f ü r oder g e g e n eine Folgeuntersuchung der Kosten treffen lassen. (In unserem

---

1) Der Name N u l l hypothese erklärt sich aus der Tatsache, daß - wie in unserem Beispiel - häufig zu entscheiden ist, ob eine Änderung des bestehenden Zustandes eingetreten ist oder nicht. Dann wird als Hypothese formuliert: Es ist keine Änderung - d. h. eine Änderung von Null - eingetreten.

2) Vgl. hierzu auch S. 70 f.

3) Vgl. aber S. 182 ff.

Beispiel träte eine solche Konfliktsituation auf, da die Werte der Beobachtungsperioden 53 und 54 bei einem Sicherheitsgrad von 99% noch gerade i n n e r h a l b der Kontrollgrenzen liegen.)[1]

6.2.4 Ermittlung eines kritischen Wertes für den Fehler erster Art

Auf die für α zu wählende Größe wirken zwei entgegengesetzte Überlegungen:
Je höher die Kosten für eine weitergehende Untersuchung der Kostenstellenkosten bzw. Verbrauchsmengen veranschlagt werden, die man außer Kontrolle vermutet, um so niedriger wird man den Wert für α wählen wollen, weil bei niedrigem α und entsprechend hohem Konfidenzniveau weniger Werte außerhalb der Kontrollgrenzen liegen als bei einem hohen Wert für α und deshalb auch in einer geringeren Anzahl von Fällen weitere Untersuchungen der Kosten durchgeführt werden. Abb. 6.3 zeigt, daß bei höherem Wert für α die Kontrollgrenzen auf den Mittelwert zurücken und bei stetiger Vergrößerung von α und Verringerung des Sicherheitsgrades immer mehr Beobachtungswerte außerhalb der Kontrollgrenzen fallen.

Demgegenüber wird das Risiko eines hohen Verlustes, der bei Nichtuntersuchung von Kosten, die tatsächlich außer Kontrolle sind, eintreten könnte, einen Anreiz bilden, einen höheren Wert für α zu wählen.[2] Je höher der Wert

---

1) Bei einem z-Wert von z = 2,58, der einem Signifikanzniveau von 0,01 bei Anwendung der Normalverteilung entspricht, ergibt sich die obere Kontrollgrenze folgendermaßen: 245,9 + 2,58 • 5,5 = 245,9 + 14,2 = 260,1.

2) Vgl. R.E. JENSEN: a.a.O., S. 269; hier wird jedoch diese Beziehung falsch wiedergegeben.

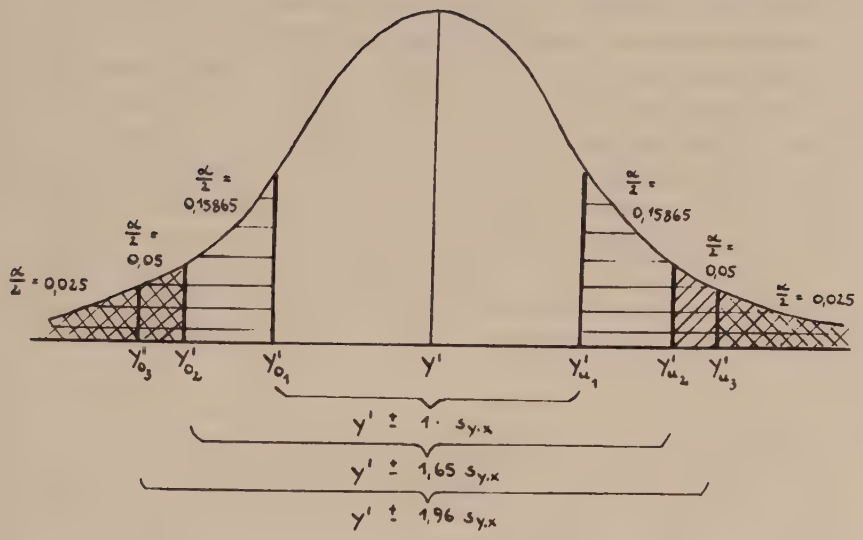


Abb. 6.3: Kontrollgrenzen bei verschiedenen Werten für α (Signifikanzniveau).

für α angesetzt wird, und je geringer entsprechend das Konfidenzniveau ist, das nach Meinung oder Erfahrung des Untersuchenden ausreichend ist, um die Kosten "außer Kontrolle" zu vermuten, um so eher werden festgestellte Abweichungen der tatsächlichen von den geschätzten Kosten außerhalb der Kontrollgrenzen liegen und um so eher wird man auf Grund dieser Abweichungen eine weitergehende Analyse der Kosten oder Verbrauchsmengen vornehmen. Bei niedrigem α und entsprechend hohem Konfidenzniveau müssen schon ganz erhebliche Abweichungen zwischen den durch die Regressionsfunktion geschätzten und den tatsächlich beobachteten Kosten vorliegen, um eine neue Kostenuntersuchung zu veranlassen.

Wie erwähnt, beruht die Entscheidung für eine bestimmte Höhe von α weitgehend auf subjektiven Schätzungen über die Höhe der Kosten einer weiteren Untersuchung der Kostenstellenkosten einerseits und der Höhe eines eventuellen Verlustes bei Nichtuntersuchung von außer Kontrolle liegenden Kosten andererseits. Allerdings kann es in bestimmten Fällen sogar möglich sein, dieses subjektive Urteil durch die Berechnung eines kritischen Wertes für α zu ersetzen, vorausgesetzt, die Kosten für eine eingehende Analyse und die möglichen Verluste aus einer nicht durchgeführten Untersuchung sind bekannt und lassen sich in quantitativer Form ausdrücken.[1)]

Der folgende Lösungsansatz soll dieses Problem verdeutlichen.[2)] Man geht zunächst von der trivialen Feststellung aus, daß die Wahrscheinlichkeit, daß eine Abweichung entweder durch den Zufall oder durch andere Ursachen erklärt werden kann, d. h. daß sie irgendeine Ursache hat, gleich 100% oder gleich 1,00 ist. Die Wahrscheinlichkeit, daß die beobachtete oder eine größere Abweichung vom Regreßwert zufallsabhängig ist, betrage α, nämlich das Risiko eines Fehlers erster Art. Demnach ist die Wahrscheinlichkeit dafür, daß diese Abweichung nicht vom Zufall abhängt, sondern andere, im Produktionsprozeß liegende Ursachen hat, gleich $(1 - \alpha)$ oder $(100\% - \alpha\%)$.

---

1) Vgl. auch R.E. JENSEN, a.a.O., S. 269.

2) Die folgenden Ausführungen sind angelehnt an die Darstellung von H.-M. SCHÖNFELD: Die Anwendung ..., a.a.O., (1968), S. 252 ff. Die Entscheidungsregel wurde von H. BIERMAN Jr. in seinem Buch: Topics in Cost Accounting and Decision, New York 1963, entwickelt. Zitiert bei H.-M. SCHÖNFELD: Die Anwendung ..., a.a.O., (1968), S. 252 ff. Dem Verfasser erscheint der Hinweis wichtig, daß es sich bei der folgenden Darstellung nicht um ein exaktes analytisches Verfahren, sondern um eine Art Probiermethode handelt, die allerdings immer dann zu brauchbaren Ergebnissen führt, wenn man auf Grund von Erfahrungswerten "vernünftige" Hypothesen ohne Extremwerte bildet.

Bezeichnen wir jetzt noch die Kosten einer weitergehenden Untersuchung der Abweichungen mit "K", die möglichen Verluste, die durch Unterlassung der Untersuchung auftreten können,[1] mit "V", und die Wahrscheinlichkeit ihres Auftretens mit "W", so erhalten wir die folgenden Tabellen:

Ursache der Abweichung	W	K	erwartete K
Zufall	α	K	$K \cdot \alpha$
andere Ursachen	$1 - \alpha$	K	$K(1 - \alpha)$
Σ	1	·	K

Tabelle 6.2: Ermittlung der erwarteten Gesamtkosten im Zusammenhang mit der Festlegung eines kritischen Wertes für α (Signifikanzniveau).

Ursache der Abweichung	W	V	erwartete V
Zufall	α	0	0
andere Ursachen	$1 - \alpha$	V	$V(1 - \alpha)$
Σ	1	·	$V(1 - \alpha)$

Tabelle 6.3: Ermittlung der erwarteten Gesamtverluste im Zusammenhang mit der Festlegung eines kritischen Wertes für α (Signifikanzniveau).

1) "V ist wie folgt berechnet: Die Gesamtabweichung kann als Verlust angesehen werden; diese Abweichung kann sich mehrmals in der Zukunft wiederholen (...); dieser Gesamtwert ist zu schätzen und auf den Gegenwartswert zu diskontieren. Auf diese Weise kann der maximale Wert der zukünftigen Verluste bei Unterlassung einer Untersuchung dieser Kostenabweichung ermittelt werden." Vgl. H.-M. SCHÖNFELD: Die Anwendung ..., a.a.O., (1968), S. 252.

Anhand dieses Schemas können wir folgende Aussagen treffen:[1)]

1. Wenn gilt: $K > V(1 - \alpha)$,
   ist eine weitergehende Untersuchung der Kosten nicht sinnvoll.

2. Wenn gilt: $K = V(1 - \alpha)$,
   kann eine Entscheidung für oder gegen eine Weiteruntersuchung der Kosten auf Grund unseres Entscheidungskriteriums nicht gefällt werden. Es handelt sich um einen indifferenten Grenzwert.

3. Wenn gilt: $K < V(1 - \alpha)$,
   sollte eine Untersuchung der Kostenhöhe vorgenommen werden.

Wählt man für K und α verschiedene Werte, so kann man eine Indifferenzkurvenschar ermitteln, von denen jede Kurve für verschiedene Werte von α die Konstellation angibt, bei denen Kosten und erwartete Verluste sich ausgleichen. Man löst dazu die Gleichung $K = V(1 - \alpha)$ nach V auf und erhält:

$$V = \frac{K}{1 - \alpha} . \qquad (6.3)$$

Wählen wir für K zum Beispiel den Wert K = 100, so erhalten wir für verschiedene Werte von α folgende Werte für V, die alle auf einer Indifferenzkurve liegen:

---

1) Vgl. H.-M. SCHÖNFELD: Die Anwendung ..., a.a.O., (1968), S. 253.

α	V
0,01	101,01
0,02	102,04
0,05	105,26
0,10	111,11
0,20	125,00
0,60	250,00
0,80	500,00
0,90	1000,00

Tabelle 6.4: Ermittlung der den Kosten K = 100 indifferenten Verluste V bei unterschiedlichem Signifikanzniveau α .

Der Wert für V_1 = 101,01 bei einem α von 0,01 sagt dabei aus, daß bei Kosten für die Untersuchung in Höhe von K = 100 und Verlusten aus einer Nichtuntersuchung in Höhe von V = 101,01 der kritische Wert für α bei 0,01 liegt. Will man also seiner Entscheidung für oder gegen eine weitere Untersuchung der Kosten eine Wahrscheinlichkeit von 99% zugrunde legen, dann sind bei Untersuchungskosten in Höhe von 100 Geldeinheiten weitere Kostenanalysen durchzuführen, wenn die erwarteten Verluste höher sind als 101,01 Geldeinheiten.

Wählt man für K weitere Werte, zum Beispiel die Werte 200, 300, 400, 500, 600, 700 und 800, so ergibt sich die in Abb. 6.4 wiedergegebene Indifferenzkurvenschar.

Anhand der Indifferenzkurve für K = 100 kann man zum Beispiel folgende Aussage treffen:
Wenn die Kosten für eine Untersuchung mit 100 Geldeinheiten und die erwarteten Verluste durch eine Nichtuntersuchung der Kostenabweichungen - wenn der Prozeß tatsächlich außer Kontrolle ist - mit 500 Geldeinheiten veranschlagt werden, ist α = 0,80 der kritische Wert. Das be-

deutet, nur wenn man bereit ist, ein Fehlerrisiko von 80% einzugehen, kann man bei den gegebenen Daten eine weitere Untersuchung unterlassen. Bei jedem geringeren Fehlerrisiko muß man zu der Entscheidung gelangen, daß eine Untersuchung der Kostenabweichungen erforderlich ist, da die erwarteten Verluste bei weitem die Kosten für diese Analyse überwiegen.

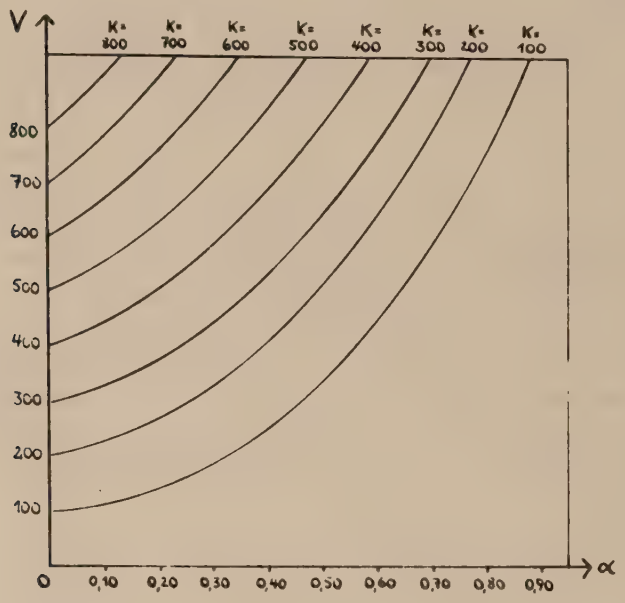


Abb. 6.4: Indifferenzkurven für Kosten K und Verluste V bei verschiedenen Werten von α .

Auf weitere Verfeinerungen dieses Modells soll hier nicht näher eingegangen werden, da sie die Problematik der Regressions- und Korrelationsanalyse unmittelbar nicht mehr berühren. Für interessierte Leser mag der Hinweis genügen,

daß SCHÖNFELD zum Beispiel zeigt, wie die Kontrolle durch die ständige Entnahme von Stichproben aus dem laufenden Produktionsprozeß erfolgt, wobei es möglich ist, in Abhängigkeit von der Höhe des maximal zulässigen Zufallsfehlers einen optimalen Stichprobenumfang festzulegen.[1] Im übrigen ist diese Vorgehensweise mit der hier beschriebenen identisch.

6.2.5 Der "Signaleffekt" der Kontrollkartentechnik

Die in den vorhergehenden Abschnitten aufgezeigte Möglichkeit der Bildung von Kostenkontrollkarten weist starke Gemeinsamkeiten mit der in der Praxis schon weitgehend verwandten Kontrollkartentechnik für statistische Qualitätskontrollen auf. Auch anhand der Kostenkontrollkarten lassen sich keine Aussagen darüber treffen, w a r u m die Kosten zu einem bestimmten Beobachtungszeitraum ungewöhnlich hoch oder niedrig sind. Sie geben also keinen Anhaltspunkt dafür, aus welchen Gründen außergewöhnlich hohe Kosten oder Verbrauchsmengen entstanden sind und wie sie verhindert werden können.[2] Durch die einbezogenen Elemente der Wahrscheinlichkeitsrechnung werden auf diese Weise jedoch die beobachteten Abweichungen in lediglich zufallsbedingte und in betrieblich bedingte Abweichungen unterteilt. Hierdurch kann die weitere Untersuchung der Kosten auf wirklich beeinflußbare Abweichungen beschränkt werden. Durch ihren "Signaleffekt" ermöglicht die Technik der Kostenkontrollkarte die Verwirklichung der Konzeption des "management by exception",[3] weil nur bei "erheblichen"

1) Vgl. H.-M. SCHÖNFELD: Die Anwendung ..., a.a.O., (1969), S. 22 ff.

2) Vgl. auch E. MANSFIELD u. H.H. WEIN, a.a.O., S. 56.

3) Vgl. auch E.E. COMISKEY, a.a.O., S. 237.; P.R. MC CLENON, a.a.O., S. 543.

Abweichungen[1] die Unternehmensleitung in den bisherigen Ablauf des Produktionsprozesses eingreift.

6.3 Zwischenbetrieblicher Kostenvergleich

6.3.1 Einführung

Ein weiteres Mittel der Betriebskontrolle bildet der zwischenbetriebliche Kostenvergleich, bei dem die Kosten verschiedener, aber verwandter Betriebe, die in den gleichen Perioden beobachtet wurden, einander gegenübergestellt werden.[2] Die Anwendung der Regressionsanalyse und der damit verbundenen statistischen Testverfahren ist also keineswegs auf Z e i t v e r g l e i c h e von Kosten im Rahmen des industriellen Rechnungswesens beschränkt.

Beispiele für Möglichkeiten zu zwischenbetrieblichen Kostenvergleichen lassen sich in großer Zahl angeben. Besonders in Staaten, in denen das sogenannte "Prinzip plandeterminierter Leistungserstellung" verwirklicht wird,[3] werden häufig die Kosten verwandter Betriebe miteinander verglichen, um Anhaltspunkte über die Arbeitsweise in diesen Betrieben zu erhalten. Aber auch in westlichen Industrieländern können zwischenbetriebliche Kostenvergleiche zur Überprüfung der Kostenhöhe mehrerer Betriebe oder Betriebsabteilungen durchgeführt werden. Man denke zum Beispiel an Unternehmen mit dezentralisierten Verkaufsbüros, deren Kosten mittels Regressionsanalyse überwacht werden. Hier lassen sich die Kosten dieser Büros während derselben Abrechnungsperiode

---

1) wobei der Begriff "erheblich" durch die Festlegung der Kontrollgrenzen konkretisiert wird.

2) E. SCHMALENBACH, a.a.O., S. 435.

3) Zur Definition des Begriffs "plandeterminierte Leistungserstellung" vgl. E. GUTENBERG: Einführung in die Betriebswirtschaftslehre. (Die Wirtschaftswissenschaften. Reihe A, Betriebswirtschaftslehre, Nr. 1, Hrsg. E. GUTENBERG.) Wiesbaden 1958. S. 190 f.

vergleichen, so daß sofort festzustellen ist, in welchem Büro die Kosten vom normalen Maß abweichen. Eine wichtige Voraussetzung ist hierbei allerdings, daß genügend viele Büros vorhanden sind (nicht zu kleiner Stichprobenumfang) und daß sie gleichartig aufgebaut sind und gleichartig arbeiten, mit anderen Worten, es muß eine möglichst große Anzahl möglichst homogener Untersuchungseinheiten vorliegen.[1] In diesem Fall hätte man eine Art von zwischenbetrieblichem Kostenvergleich vorgenommen, wenn man die verschiedenen Verkaufsbüros als "selbständige Betriebe" ansieht.

Im folgenden Abschnitt soll anhand eines Zahlenbeispiels gezeigt werden, auf welche Weise eine n i c h t l i n e a r e Regressionsanalyse Aufschlüsse über die Kostenhöhe mehrerer Betriebe geben kann, wenn die Stückkosten dieser Betriebe in Abhängigkeit vom Produktionsvolumen als nichtlineare Funktionen dargestellt werden.[2]

6.3.2 Stückkostenuntersuchung mehrerer Betriebe unter Verwendung nichtlinearer Regressionsfunktionen

Bereits im 2. Kapitel bei der Darstellung der Grundlagen der Regressions- und Korrelationsanalyse wurde darauf hingewiesen, daß in aller Regel lineare Funktionsformen für

1) Vgl. auch E.E. COMISKEY, a.a.O., S. 235, und den Abschnitt 7.3 dieser Arbeit.

2) Die Beschränkung in der Darstellung nichtlinearer Regressionen auf ein Beispiel aus dem Bereich zwischenbetrieblicher Kostenvergleiche erschien dem Verfasser sinnvoll, da hierdurch die Problematik nichtlinearer Gleichungsansätze besonders deutlich zum Ausdruck kommt. Gleichzeitig soll jedoch darauf hingewiesen werden, daß das Problem nichtlinearer regressionsanalytischer Kostenuntersuchungen keinesfalls auf zwischenbetriebliche Kostenvergleiche beschränkt ist, sondern grundsätzlich bei allen Anwendungsmöglichkeiten auftritt. Vgl. auch S. 20 ff. und S. 200 ff.

Kostenuntersuchungen genügend exakt sind und man sich auch aus Gründen der einfacheren Berechnungsmethode für die Regressions- und Korrelationsparameter auf lineare Ansätze beschränkt.[1] Dennoch treten in der betrieblichen Kostenrechnung Relationen auf, die nicht durch lineare Regressionsfunktionen beschrieben werden können und sich auch nicht für ein bestimmtes Intervall hinreichend genau durch eine Gerade approximieren lassen. In diesen Fällen steht der Analysierende vor der Aufgabe, aus der großen Zahl möglicher Gleichungsansätze den "richtigen" oder zumindest einen "geeigneten" Ansatz auszuwählen. Dabei kann er den Verlauf der Punktwolke im Koordinatensystem und Erkenntnisse aus früheren Untersuchungen oder theoretischen Überlegungen als Anhaltspunkte für seine Entscheidung heranziehen.[2]

Neben dem Gesichtspunkt des Rechenaufwands, der für die Wahl einer möglichst einfachen Kurvenform spricht,[3] sollte man vor Durchführung der Analyse stets die Frage stellen, was mit der errechneten Regressionsfunktion beabsichtigt wird. Sollen auf Grund der Regreßwerte zukünftige Schätzungen nur in dem beobachteten Bereich vorgenommen werden oder sind auch über diesen Bereich hinaus Extrapolationen geplant?[4] Soll die Regressionslinie das b e o b a c h t e t e Datenmaterial möglichst exakt wiedergeben oder soll sie vielmehr theoretischen Erkenntnissen entsprechen, auch wenn sie in einem konkreten Fall vielleicht für den Bereich, in dem Beobachtungswerte vorliegen und der ja nur ein Ausschnitt aus dem gesamten möglichen Untersuchungsbereich sein kann,[5] nicht so exakte

---

1) Vgl. hierzu auch die Darstellungen im Abschnitt 6.3.3.

2) Vgl. S. 20.

3) Vgl. aber S. 211 ff., sowie 9. Anhang I, wo auf die Berechnung von Polynomen mit elektronischen Datenverarbeitungsanlagen eingegangen wird.

4) Vgl. dazu aber S. 201 ff.

5) Vgl. auch S. 144.

Schätzwerte liefert wie eine andere Funktionsform, die wiederum - verwendet man sie über den Beobachtungsbereich hinaus - zu unsinnigen Ergebnissen führt?

Mit Hilfe des folgenden Zahlenbeispiels soll versucht werden, Antworten auf diese Fragen zu geben. Wir gehen von der Annahme aus, daß wir in 15 Betrieben während eines bestimmten Zeitraumes die Selbstkosten je Stück[1] (y_i) und die mit diesen Kosten erzeugte Stückzahl an identischen Produkten (x_i, wobei diese Angaben in 1000 Stück erfolgen) beobachtet und folgende Werte erhalten haben:[2]

Betrieb	y_i	x_i
1	8	2
2	10	3
3	7	4
4	6	4
5	5	5
6	5	6
7	4	6
8	3	6
9	4	7
10	5	8
11	3	9
12	2	10
13	1	12
14	1	13
15	2	14
$\sum$	66	109

Tabelle 6.5: Selbstkosten je Stück (y_i) und produzierte Stückzahl (x_i) in 15 Betrieben.

1) Da es sich bei den Stückkosten um statistische Verhältniszahlen handelt, treten bei der Korrelation besondere Probleme auf. Im Rahmen dieser Arbeit kann hierauf jedoch nicht weiter eingegangen werden. Vgl. E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 201 ff.

2) Entnommen aus E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 173.

Aus der Darstellung der Punkte im Koordinatensystem geht hervor (vgl. Abb. 6.5), daß eine lineare Regressionsfunktion in diesem Fall auf Grund der eindeutig nichtlinearen Tendenz nicht in Erwägung gezogen zu werden braucht. Auch die Kostentheorie lehrt, daß die gesamten Stückkosten nicht einer linearen Funktion folgen, sondern einen, wie auch immer aussehenden, gekrümmten Verlauf nehmen.

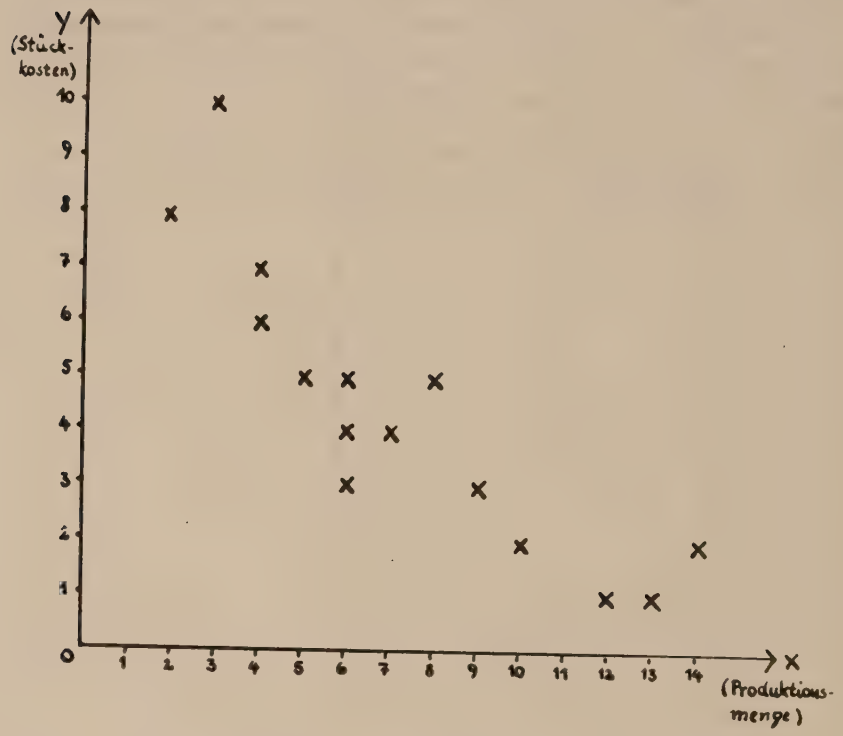


Abb. 6.5: Stückkosten (y) in Abhängigkeit von der Produnktionsmenge (x) für 15 Betriebe.

Wir wollen deshalb zunächst eine ganze rationale Funktion zweiten Grades (Parabel) untersuchen. Nach der im Abschnitt 2.3 dargestellten Berechnungsmethode für nichtlineare Re-

gressionsfunktionen erhalten wir auf Grund des Datenmaterials der Tabelle 6.5 die Regressionsgleichung[1)]

$$y' = 11{,}87 - 1{,}541\ x + 0{,}0571\ x^2,$$

und die in Abb. 6.6 gestrichelt eingezeichnete Regressionskurve.

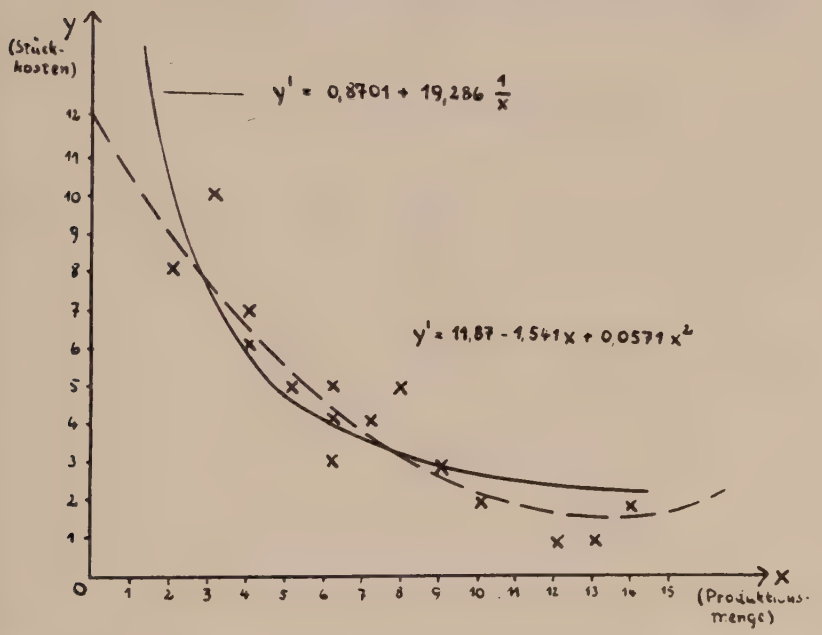


Abb. 6.6: Regressionsfunktionen für die Abhängigkeit der Stückkosten (y) vom Produktionsvolumen (x).

1) Die Werte wurden unverändert übernommen von E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 174.

Aus der Darstellung dieser Kurve und der Tabelle 6.6 der Regreßwerte y'_i erkennt man, daß bei einer Produktionsmenge von x = 15 die mit Hilfe der Regressionsfunktion ermittelten Stückkosten y' größer sind als die Stückkosten, die sich bei einer Produktion in Höhe von 14 Einheiten (x = 14) ergeben. Bei einer Produktion von x = 0 weisen die Stückkosten einen Wert von y'_o = 11,87 Geldeinheiten auf. Zumindest das Ergebnis, daß die Stückkosten 11,87 Geldeinheiten betragen, wenn überhaupt nicht produziert wird, ist vom ökonomischen Standpunkt aus nicht zu erklären.[1)]

Betrieb	y'_i	x_i
1	9,02	2
2	7,76	3
3	6,62	4
4	6,62	4
5	5,59	5
6	4,68	6
7	4,68	6
8	4,68	6
9	3,88	7
10	3,18	8
11	2,63	9
12	2,17	10
13	1,60	12
14	1,49	13
15	1,49	14

Tabelle 6.6: Produktionszahlen (x_i) und zugehörige Regreßwerte (y'_i) bei Verwendung einer Parabel der Form $y' = a - bx + cx^2$ als Regressionsfunktion.

1) Der Wiederanstieg der Stückkosten läßt sich immerhin mit den von einem bestimmten Punkt an auftretenden Kosten der Überbeschäftigung erklären, wenn die Gesamtkostenkurve bis zu diesem Punkt linear verläuft und dann überproportional ansteigt. Vgl. auch Abb. 5.2 auf S. 90.

Aus diesem Grund erscheint es sinnvoll, einen anderen Gleichungsansatz zu wählen. Dafür bietet sich die Hyperbelfunktion[1)]

$$y' = a_{yx} + b_{yx} \frac{1}{x}$$

an, die asymptotisch zur Ordinate verläuft, wenn man von einer linearen Gesamtkostenkurve ausgeht. In diesem Fall ergeben die Zahlenwerte der Tabelle 6.5 die Regressionsfunktion[2)]

$$y' = 0{,}8701 + 19{,}286 \frac{1}{x} .$$

Sie wird durch die nichtunterbrochene Linie in der Abb. 6.6 wiedergegeben. Die auf Grund dieser Funktion ermittelten Regreßwerte sind in Tabelle 6.7 aufgeführt.

Die graphische Darstellung dieser Funktion und die Regreßwerte lassen erkennen, daß die Stückkosten mit zunehmender Ausbringung ständig abnehmen und sich dem Wert 0,8701 nähern, während bei einer Produktion von x = 0 sich für die Stückkosten der Schätzwert $y'_0 = \infty$ ergibt. Dieser Wert ist zwar nur ein fiktiver Rechenwert, da es ökonomisch sinnlos ist, bei einer Produktion von null die Gesamtkosten durch null zu dividieren, aber die Beziehung

$$y' = a_{yx} + b_{yx} \frac{1}{x}$$

---

1) Zur praktischen Berechnung von Hyperbelfunktionen vgl. I. CHAWIN u. W. KORSHEWITSCH: Die Anwendung mathematischer Methoden und elektronischer Rechenmaschinen für die Analyse der Selbstkosten. "Statistische Praxis. Zeitschrift für Statistik und Rechnungswesen". 19 (1964) 4, S. 107 ff.

2) Die Werte wurden übernommen von E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 175.

Betrieb	y'_i	x_i
1	10,51	2
2	7,30	3
3	5,69	4
4	5,69	4
5	4,73	5
6	4,09	6
7	4,09	6
8	4,09	6
9	3,63	7
10	3,28	8
11	3,01	9
12	2,80	10
13	2,48	12
14	2,35	13
15	2,25	14

Tabelle 6.7: Produktionszahlen (x_i) und zugehörige Regreßwerte (y'_i) bei Verwendung einer Hyperbel der Form $y' = a + b \frac{1}{x}$ als Regressionsfunktion.

gibt die Degression der fixen Stückkosten bei fortlaufender Steigerung der Produktion weitaus besser wieder als die ursprünglich gewählte Beziehung

$$y' = a_{yx} + b_{yx}\, x + c_{yx} x^2.$$

Sie eignet sich damit besser zur a l l g e m e i n e n Darstellung der Abhängigkeit der Stückkosten von der Produktionsmenge.[1] Trotzdem kann für den beobachteten Produktionsbereich ($x = 2$ bis $x = 14$) die Parabel

$$y' = 11{,}87 - 1{,}541\, x + 0{,}0571\, x^2$$

1) Vgl. E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 175.

zu besseren Schätzwerten führen als die theoretisch exaktere Hyperbel

$$y' = 0{,}8701 + 19{,}286 \frac{1}{x} .$$

Diese Behauptung läßt sich mit Hilfe der Standardfehler der Schätzung für jede Regressionsfunktion beweisen. Im Abschnitt 5.6.4 und auch im Abschnitt 6.2.1 wurde dargestellt, daß der Standardfehler der Schätzung $\sigma_{y.x}$ ein Maß für die Streuung der Residuen um die Regressionslinie ist und daß damit seine Größe Aufschluß über die Anpassung der Regressionsfunktion an die empirischen Beobachtungswerte gibt. Je größer der Wert für $\sigma_{y.x}$ ist, um so größer ist die Streuung und um so schlechter die Anpassung.

Für die Parabel ergibt sich ein Schätzwert für den Standardfehler der Schätzung von ${}_P\hat{\sigma}_{y.x} = 1{,}106$ und für die Hyperbel von ${}_H\hat{\sigma}_{y.x} = 1{,}395$, woraus geschlossen werden kann, daß die Anpassung bei Verwendung der Parabel besser gelingt.[1] In gleicher Weise geben die Korrelationskoeffizienten bzw. die Bestimmtheitsmaße Auskunft über die Güte der Anpassung. Einem ${}_Pr_{yx} = 0{,}922$ der Parabel steht ein ${}_Hr_{yx} = 0{,}854$ der Hyperbel gegenüber.[2] Diese Größen bestätigen die anhand der Standardfehler der Schätzung getroffenen Aussagen.

Das Fazit dieser Untersuchung lautet damit, daß i m g e g e b e n e n B e r e i c h die Parabel besser für Schätzzwecke geeignet ist als die Hyperbel, obwohl sie Werte liefert, die ökonomisch schwer zu interpretieren sind. "Die Hyperbelfunktion dagegen ist in unserem Beispiel, ..., für eine allgemeine theoretisch-ökonomische Darstellung vorzuziehen".[3]

---

1) Auf die Berechnung von Konfidenzintervallen wird hier verzichtet. Vgl. dazu den Abschnitt 6.2.1.

2) Die Zahlen wurden übernommen von E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 183.

3) E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 183.

6.3.3 Die Form der Regressionsfunktion

6.3.3.1 Allgemeine Überlegungen

Auf die Probleme, die für den Kostenrechner bei der Wahl der geeigneten Funktionsform auftreten, wurde schon kurz an mehreren Stellen dieser Arbeit eingegangen. Wegen der großen Bedeutung dieses Punktes sollen in diesem Abschnitt die wichtigsten Gedanken noch einmal zusammengefaßt und vertieft werden.

Bei den im Rahmen dieser Arbeit durchgerechneten oder nur erwähnten Beispielen wurde fast ausschließlich von linearen Regressionsfunktionen ausgegangen. Dafür sprach einmal die Tatsache, daß die moderne ökonomische Theorie weitgehend lineare Kostenfunktionen zugrundelegt,[1] zum anderen, daß - wie bei der Ermittlung der Sollkostenkurve in der Plankostenrechnung oder der Bestimmung der optimalen Schlüsselgröße für die Gemeinkostenverrechnung - von der Aufgabenstellung her nur geradlinige Regressionsfunktionen in Betracht zu ziehen waren. Darüber hinaus lehrt die Erfahrung, daß sich auch gekrümmte Regressionslinien für einen T e i l b e r e i c h recht gut durch lineare Funktionen darstellen lassen.[2] Der für die Aussagefähigkeit der Regressionsergebnisse nachteilige, aber gerade für empirische Kostenuntersuchungen typische relativ geringe Streubereich in den Werten der unabhängigen Variablen[3] wirkt sich damit insoweit auf die Analyse aus, als er die Wahl der Funktionsform in Richtung auf lineare

1) Zumindest in Bereichen normaler Beschäftigung. Vgl. z. B. E. GUTENBERG, a.a.O., (1970), S. 336 ff. Vgl. aber auch das 3. Kapitel dieser Arbeit.

2) Vgl. S. 52 ff. und den Abschnitt 6.3.3.2.

3) Vgl. den Abschnitt 7.3.

Gleichungsansätze erleichtert bzw. fördert. Auch der gerade für die Analyse von Kostenarten oder ihrer Mengenkomponenten zu begrüßende Trend von der einfachen zur multiplen Regressionsrechnung führt zu einer weitgehenden Vernachlässigung von nichtlinearen Funktionstypen. Während nämlich bei einfachen Analysen häufig noch die Form der in ein Streupunktdiagramm eingezeichneten Punktwolke Aufschlüsse über eine spezielle Regressionslinie gibt,[1] ist man bei multiplen Analysen mit mehr als zwei unabhängigen Veränderlichen weitgehend auf theoretische Kenntnisse oder Überlegungen angewiesen. Gerade in Bezug auf die Form von Kostenfunktionen sprechen diese aber eindeutig für einen linearen Ansatz.

6.3.3.2 Extrapolation der Regressionsfunktion bei Approximation einer Kurve durch eine Gerade

Es wurde bereits in früheren Abschnitten gezeigt, daß sich für einen bestimmten, begrenzten Teilbereich, in dem Beobachtungswerte für die unabhängige Variable vorliegen, eine als theoretisch richtig angesehene gekrümmte Regressionslinie relativ gut durch eine Gerade approximieren läßt.[2] Für x-Werte, die innerhalb dieses Intervalls liegen, gelangt man auf Grund der linearen Regressionsfunktion in der Regel zu praktisch ausreichend genauen Schätzwerten für die abhängige Variable. Problematisch wird dieses Verfahren jedoch, wenn Extrapolationen vorgenommen werden, d. h. wenn auf Grund zukünftiger Beobachtungswerte für die unabhängige Variable, die größer oder kleiner sind als die in die Analyse aufgenommenen Werte,

---

1) Vgl. hierzu aber den Abschnitt 6.3.3.4.

2) Vgl. z. B. die Abb. 4.2 auf S. 54 und insbesondere den Abschnitt 6.3.3.3.2.

Schätzwerte für die Kosten (y'_i-Werte) ermittelt werden sollen. Für den Fall der Approximation einer gekrümmten Regressionslinie durch eine Gerade in einem Teilbereich gilt strenger noch als für die auch theoretisch richtige Ermittlung einer linearen Funktion, daß die mit Hilfe der Regressionsanalyse errechneten Schätzwerte nur für das Intervall Bedeutung haben, das die beobachteten Werte der unabhängigen Variablen umschließen.[1] Eine Verlängerung der Regressionslinie bis zum Ursprung oder in Richtung auf höhere Werte der unabhängigen Veränderlichen ist in diesen Fällen grundsätzlich unzulässig.[2] Hier sprechen nämlich logische Gründe dafür, daß die durch die lineare Regressionsfunktion ausgedrückte Relation außerhalb des Beobachtungsraumes keine Gültigkeit haben kann.[3] Man begibt sich durch die Extrapolation auf ein Gebiet, das infolge des fehlenden Datenmaterials nicht untersucht werden konnte, von dem man aber auf Grund anderer Kenntnisse weiß, daß hier die Relation zwischen y und x durch eine Gerade auch nicht mehr annähernd exakt wiedergegeben werden kann und daß die lineare Regressionsfunktion in diesem Bereich damit zu unsinnigen Schätzwerten führen muß.[4] Schließlich sei noch darauf hingewiesen, daß in manchen Fällen nichtlineare Funktionstypen auf lineare Formen zurückgeführt werden können, so zum Beispiel durch Logarithmierung der Funktion oder durch Umwandlung einer Regression mit einer unabhän-

---

1) Vgl. M. EZEKIEL u. K.A. FOX, a.a.O., S. 65 u. S. 115 f.

2) Vgl. auch D.G. BEECH: Experiences of Correlation Analysis. "Applied Statistics. A Journal of the Royal Statistical Society". 2 (1953) 2, S. 73 ff., hier S. 74.; P.G. MOORE: Regression as an Analytical Tool. "Applied Statistics. A Journal of the Royal Statistical Society". 11 (1962) 2, S. 106 ff., hier S. 106 f.; W.A. WALLIS u. H.V. ROBERTS, a.a.O., S. 455.

3) Vgl. M. EZEKIEL u. K.A. FOX, a.a.O., S. 65. F. EGERMAYER: Die Korrelationsanalyse ..., a.a.O., S. 436.

4) Vgl. auch das Beispiel im Abschnitt 6.3.3.3.2.

gigen Variablen, die auch als quadratisches Glied in der Gleichung erscheint (Parabel)[1], in eine multiple Regression, wobei das quadratische Glied als zweite unabhängige Veränderliche der mehrfachen Regression aufgefaßt wird.[2]

6.3.3.3 Prüfung der Linearität einer Regression

In der Statistik sind verschiedene Verfahren bekannt, die es ermöglichen, in einem bestimmten Fall zu testen, ob für das vorliegende Datenmaterial eine lineare oder eine nichtlineare Regressionsfunktion angemessen ist. Diese Testverfahren sind insbesondere dann empfehlenswert, wenn keinerlei andere Informationen - über die Form der Relation zwischen abhängiger und unabhängiger Variabler vorliegen als das empirische Beobachtungsmaterial.

6.3.3.3.1 Muster der Residuen

Eine dieser Möglichkeiten besteht darin, zunächst eine vorläufige lineare Regressionsanalyse durchzuführen und dann die Residuen ($y_i - y'_i$) in ein Koordinatenkreuz einzuzeichnen und nach Mustern ("residual patterns")[2] d.h. nach einem bestimmten Verteilungsverlauf dieser Residuen zu suchen. Residuen können für die Linearitätsprüfung einer Regression oft nützlicher als der Korrelationskoeffizient oder das Bestimmtheitsmaß sein, da sich auch bei hohem Korrelationskoeffizienten - der ja als Indiz für die Richtigkeit der verwendeten Funktionsform angesehen werden kann - Muster in den Residuen zeigen können.[4]

---

1) Die Regressionsfunktion lautet: $y' = a + bx + cx^2$.

2) Vgl. E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 178.; A.S. GOLDBERGER, a.a.O., S. 109 ff.; aber auch R.E. JENSEN, a.a.O., S. 272.

3) A.S. GOLDBERGER, a.a.O., S. 130 f.

4) Ebenda. S. 131.

Zeichnet sich zum Beispiel ein Muster der Art ab, daß für niedrige und hohe Werte der unabhängigen Veränderlichen niedrige Residuenwerte, für mittlere Werte der unabhängigen Variablen aber hohe Werte der Residuen auftreten - vgl. die Abb. 6.7 - so sollte man die Aufnahme eines quadratischen Gliedes in die Regressionsgleichung in Erwägung ziehen.[1)]

Abb. 6.7: Möglicher Verlauf der Residuen bei Anwendung einer linearen Regressionsfunktion auf einen nichtlinearen (parabolischen) Zusammenhang.

6.3.3.3.2 Varianzanalyse

Eine weitere Möglichkeit zur Prüfung der Linearität einer Regression besteht in der Anwendung des F-Tests in Verbin-

1) Vgl. A.S. GOLDBERGER, a.a.O., S. 131 und die dort angegebene Literatur.

dung mit einer Streuungszerlegung (Varianzanalyse).[1] Die gesamte Streuung zwischen einem beobachteten Kostenbetrag y_i und dem Mittelwert $\bar{y}$ setzt sich aus zwei Abweichungen zusammen, wie aus der Abb. 6.8 entnommen werden kann:[2]

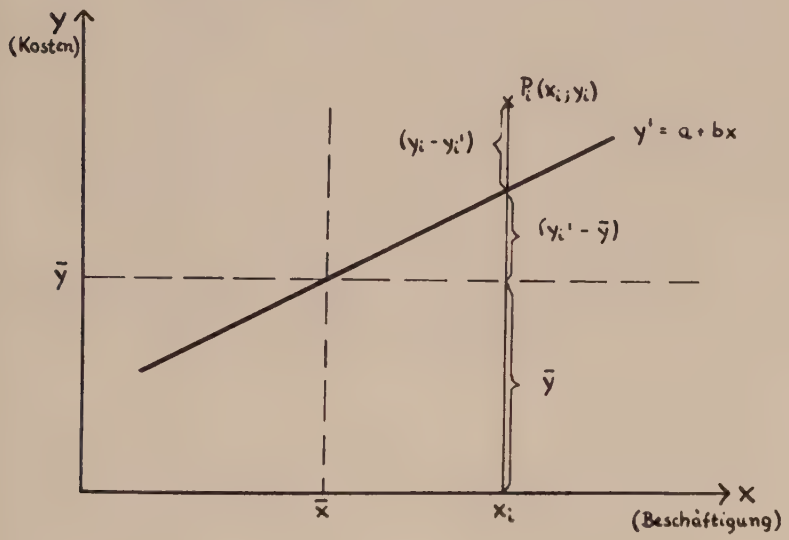


Abb. 6.8: Aufspaltung eines beobachteten Stichprobenwertes y_i als Vorbereitung für eine Varianzanalyse.

---

1) Zur ausführlichen Darstellung der Varianzanalyse vgl. insbesondere die Artikel von: G. AMERMAN: The Mathematics of Variance Analysis. "Journal of Accounting Research". 4 (1953) 3, S. 258 ff., 4 (1953) 4, S. 329 ff., 5 (1954) 1, S. 56 ff., und CH. KWANG u. A. SLAVIN: The Simple Mathematics of Variance Analysis. "The Accounting Review". 37 (1962) 3, S. 415 ff.; A. LINDER, a.a.O., S. 186 ff. M. EZEKIEL u. K.A. FOX, a.a.O., S. 405 f.

2) Vgl. B. OSTLE: Statistics in Research. Basic Concepts and Techniques for Research Workers. Second edition. Ames (Iowa, U.S.A.) 1963. S. 165 f.; Vgl. auch die ähnliche Darstellung bei F.E. CROXTON u. D.J. COWDEN, a.a.O., S. 463.

Für y_i ergibt sich aus der Aufspaltung folgende Beziehung:

$$y_i = \bar{y} + (y'_i - \bar{y}) + (y_i - y'_i). \qquad (6.4)$$

Dann gilt für die Abweichung $(y_i - \bar{y})$:

$$(y_i - \bar{y}) = (y'_i - \bar{y}) + (y_i - y'_i). \qquad (6.5)$$

Verbal ausgedrückt sagt diese Gleichung aus, daß sich die Gesamtabweichung zwischen beobachtetem Kostenwert und dem arithmetischen Mittel zusammensetzt aus der durch die Regressionsfunktion erklärten Abweichung $(y'_i - \bar{y})$ und den Residuen, d. h. den nicht erklärten Abweichungen $(y_i - y'_i)$. Es kann gezeigt werden, daß diese Beziehung ihre Gültigkeit behält, wenn man die einzelnen Abweichungen quadriert und die Summe dieser Quadrate bildet:[1)]

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y'_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - y'_i)^2. \qquad (6.6)$$

Für diese Beziehung können wir vereinfachend schreiben:

$$G = E + N,$$

wobei G ν = (n - 1), E ν = (k) und N ν = (n - k - 1) Freiheitsgrade besitzt.[2)]

Um den F-Wert zum Testen der Linearität der Regression zu ermitteln, können wir die folgende Tabelle aufstellen:[3)]

---

1) Auf diese Ableitung soll hier verzichtet werden. Vgl. z. B. K.W. SMILLIE: An Introduction to Regression and Correlation. Toronto, London und New York 1966. S. 8.

2) Vgl. Ebenda. S. 12. k = Zahl der unabhängigen Variablen in der Regressionsfunktion. Die Symbole G, E und N sollen ausdrücken, daß es sich bei G um die "Gesamtabweichung", bei E um die durch die Regressionsfunktion "Erklärte Abweichung" und bei N um die "Nicht erklärte Abweichung" handelt.

3) Vgl. z. B. M. EZEKIEL u. K.A. FOX, a.a.O., S. 406.; A. LINDER, a.a.O., S. 188 ff.; K.W. SMILLIE, a.a.O., S. 49.

Ursache (Quelle) der Abweichung	Anzahl der Freiheitsgrade	Summen der Abweichungsquadrate	Mittlere Quadrate (Durchschnittsquadrate)
Regression	k	$E = \sum_{i=1}^{n} (y'_i - \bar{y})^2$	$M_1 = \frac{E}{k}$
Residuen	$n - k - 1$	$N = \sum_{i=1}^{n} (y_i - y'_i)^2$	$M_2 = \frac{N}{n-k-1}$
Gesamt	$n - 1$	$G = \sum_{i=1}^{n} (y_i - \bar{y})^2$	•

Tabelle 6.8: Tabelle zur Varianzanalyse

Mit Hilfe der Testgröße $F = \frac{M_1}{M_2}$ kann man nun auf Grund des Stichprobenergebnisses testen, ob zwischen den Variablen der Grundgesamtheit (Y als abhängige und X als unabhängige Variable) ein linearer Zusammenhang existiert. Es handelt sich bei dieser Untersuchung um die schon ausführlich erörterte statistische Hypothesenprüfung, indem man von der Nullhypothese H_0 ausgeht, die besagt, daß zwischen den Variablen Y und X in der Gesamtheit k e i n e lineare Beziehung besteht, und diese Hypothese mit Hilfe der F-Werte testet.[1)]

Die Testgröße F folgt einer F-Verteilung[2)] mit $\nu = k$ Freiheitsgraden für M_1 und $\nu = (n - k - 1)$ Freiheitsgraden für M_2.[3)] Ist $M_1 > M_2$, d. h. ist das mittlere Quadrat der durch die Regressionsfunktion erklärten Abweichung größer als das mittlere Quadrat der nicht erklärten Streuung, so vergleicht man das auf Grund konkreten Zahlenmaterials ermittelte F mit einem vorher festzulegenden kritischen Wert $F_{\alpha}(\nu_1, \nu_2)$, wobei durch α - wie bei Anwendung der Normalverteilung - ein bestimmtes Signifikanzniveau (Sicherheitsgrad der Aussage) festgelegt wird. Ist dagegen $M_1 < M_2$, so errechnet man die Testgröße

$$F' = \frac{M_2}{M_1}$$

und vergleicht sie mit dem kritischen Wert $F_{\alpha}(\nu_2, \nu_1)$.

Um die Varianzanalyse mit einem Zahlenbeispiel weiter zu erläutern, wollen wir von dem in Tabelle 6.5 wiedergegebenen Zahlenmaterial ausgehen[4)] und für diese Beobachtungs-

---

1) Vgl. z. B. S. 69 ff.u. S. 180 ff.

2) Vgl. die Tabelle 5* der F-Verteilung im Anhang II, S. 288.

3) Vgl. K.W. SMILLIE, a.a.O., S. 49.

4) Vgl. S. 193.

werte eine l i n e a r e Einfachregression der Form $y' = a_{yx} + b_{yx}x$ ermitteln. Es sei daran erinnert, daß wir im Abschnitt 6.3.2 zwei n i c h t l i n e a r e Regressionsfunktionen - eine Parabel und eine Hyperbel - für dieses Datenmaterial aufgestellt hatten, weil sowohl der Verlauf der Punktwolke im Koordinatensystem als auch betriebswirtschaftliche Überlegungen eine lineare Form der Regression nicht sinnvoll erscheinen ließen. Wir wollen nun mit Hilfe der Varianzanalyse prüfen, ob das vorliegende Beobachtungsmaterial tatsächlich gegen die Annahme der Linearität spricht.

Für die Regressionsfunktion bei unterstellter linearer Abhängigkeit der Stückkosten von der Produktionsmenge erhalten wir die Beziehung

$$y' = 8{,}885 - 0{,}617\ x.$$

Auf Grund dieser Gleichung errechnen wir die Schätzwerte y'_i, vergleichen sie mit den beobachteten Werten y_i und erhalten für

$$\sum_{i=1}^{n} (y_i - y'_i)^2 = 21{,}64,$$

während sich für

$$\sum_{i=1}^{n} (y'_i - \bar{y})^2$$

ein Wert von 71,94 errechnet. Berücksichtigen wir den Stichprobenumfang n = 15 und die Zahl der unabhängigen Variablen (k = 1), so ergeben sich für M_1 und M_2 folgende Werte:

$$M_1 = \frac{E}{k} = \frac{\sum_{i=1}^{n} (y'_i - \bar{y})^2}{k} = \frac{71{,}94}{1} = 71{,}94 \text{ und}$$

$$M_2 = \frac{N}{n-k-1} = \frac{\sum_{i=1}^{n} (y_i - y'_i)^2}{n-k-1} = \frac{21,64}{13} = 1,67.$$

Damit lautet der auf Grund des Stichprobenmaterials ermittelte F-Wert

$$F = \frac{M_1}{M_2} = \frac{71,94}{1,67} = 43.$$

Bei Annahme eines Konfidenzniveaus von 99% (Irrtumswahrscheinlichkeit 1%) liest man aus der Tabelle der F-Verteilung einen kritischen Wert von $F_{0,99\ (\nu_1=1,\ \nu_2=13)} = 9,07$ ab.[1)] Da der errechnete F-Wert weit größer als dieser Tabellenwert ist, kommen wir zu dem - vergleicht man die Ausführungen des Abschnitts 6.3.2 - überraschenden Ergebnis, daß das in Tabelle 6.5 angegebene Datenmaterial gut durch eine lineare Regressionsfunktion dargestellt werden kann. Dieses Ergebnis ist aus der Tatsache zu erklären, daß das beobachtete Zahlenmaterial Werte für die unabhängige Variable nur in dem Bereich zwischen x = 2 und x = 14 enthält. Die "wahre", d. h. betriebswirtschaftlich sinnvolle, Beziehung zwischen Stückkosten und Produktionsvolumen hätte aber auf Grund der Stichprobe erst erkannt werden können, wenn über eine Produktionsmenge von 14 hinaus Beobachtungswerte vorgelegen hätten. In diesem Bereich verläuft die Regressionsgerade y' = 8,885 - 0,617 x schon im negativen Abschnitt des Koordinatensystems. Stückkosten in Höhe von -0,373 (bei einer Produktion von x = 15) sind jedoch vom betriebswirtschaftlichen Standpunkt aus unsinnig.

An diesem Beispiel zeigt sich wieder die schon im Abschnitt

---

1) Vgl. 10. Anhang II, Tabelle 5*, S. 288.

6.3.3.2 beschriebene Problematik der Approximation von theoretisch richtigen Kurvenverläufen durch Funktionen, die auf Grund des beobachteten Datenmaterials geeignet erscheinen. Während diese Funktionen innerhalb - aber eben nur innerhalb - des Beobachtungsbereiches der unabhängigen Variablen zu vernünftigen Schätzwerten führen, verlieren sie ihre Bedeutung, wenn Extrapolationen über diesen Bereich hinaus vorgenommen werden. Dieser Tatsache sollte man sich stets bewußt sein, wenn man auf Grund eines eng begrenzten Streuungsintervalls für die unabhängige Variable Kosten- oder Verbrauchsschätzungen mit Hilfe der Regressionsanalyse vornimmt. Insbesondere bei geringem Stichprobenumfang sollten daher immer Überlegungen in Bezug auf einen möglichen Wendepunkt, ein Maximum oder Minimum im Kurvenverlauf, negative Werte usw. der mathematischen Anpassung der Beobachtungswerte an eine bestimmte Funktionsform vorausgehen.[1)]

6.3.3.4 Zusammenfassung

Im Abschnitt 6.3.3 wurde versucht, die Problematik darzustellen, die mit der Auswahl einer für den untersuchten Zusammenhang zwischen Kosten oder Verbrauchsmengen und ihren Einflußfaktoren typischen Funktionsform verbunden ist. Wir haben gesehen, daß im allgemeinen lineare Regressionsfunktionen bevorzugt werden, da sie leichter zu berechnen sind, die Schätzungen nach der Methode der kleinsten Quadrate in diesem Fall ihre wünschenswerten Eigenschaften behalten,[2)] und sie häufig auf Grund des

1) Vgl. auch die Abschnitt 6.3.3.4 und 7.3.

2) Vgl. z. B. A.S. GOLDBERGER, a.a.O., S. 122 f.; T. YAMANE a.a.O., S. 377.; R.E. JENSEN, a.a.O., S. 272 und S. 13 dieser Arbeit.

vorliegenden begrenzten Streubereichs für die unabhängige Variable gute Approximationen für nichtlineare Formen darstellen. Jedoch besteht bei solchen Vereinfachungen die Gefahr, daß die Zuverlässigkeit der Schätzungen beeinträchtigt wird.[1] Auch hat das Argument der Rechenvereinfachung bei Verwendung linearer Gleichungsansätze in letzter Zeit durch den Einsatz programmierbarer Tischrechner und elektronischer Datenverarbeitungsanlagen an Bedeutung verloren. Programmierte Berechnungsverfahren erlauben es, ausgehend von einem linearen Ansatz über die Berechnung von Polynomen eines stets um "eins" höheren Grades eine geeignete Kurvenform zu finden.[2] Man berechnet also zunächst ein Polynom ersten Grades $y' = a + bx$ (Gerade), dann ein Polynom zweiten Grades $y' = a + bx + cx^2$, ein Polynom dritten Grades $y' = a + bx + cx^2 + dx^3$ usw. und testet dann mit Hilfe der Varianzanalyse, wie groß die Zunahme des durch die Regressionsfunktion erklärten Anteils an der Gesamtstreuung ist, der durch die Aufnahme des jeweiligen Regressionsgliedes höherer Ordnung erzielt werden konnte.[3] Wird dieser Betrag so niedrig, daß er praktisch zu keiner Verbesserung der Schätzwerte mehr führt, kann die zuletzt untersuchte Relation für Schätzzwecke als ausreichend genau angesehen werden.

Da es keine allgemeine Regel gibt, kein Rezept geben kann, das dem Analysierenden sagt, welche Funktionsform in einem konkreten Fall zu wählen ist, sollten neben den Anhaltspunkten, die das vorliegende Datenmaterial liefert, stets

---

1) Vgl. auch E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 104.

2) Auf die Berechnung von Polynomen kann hier nicht ausführlicher eingegangen werden. Vgl. z. B. F.E. CROXTON u. D.J. COWDEN, a.a.O., S. 282 u. S. 486 ff.; C.H. GOULDEN, a.a.O., S. 181 ff., und den Anhang I über die rechnerische Durchführung von Regressions- und Korrelationsanalysen, S. 252 ff.

3) C. H. GOULDEN, a.a.O., S. 181.

Überlegungen bezüglich des theoretisch möglichen Kurvenverlaufs den Ausgangspunkt der Analyse bilden. GOLDBERGER hat in anschaulicher Weise dargestellt, daß in der Regel weder die ökonomische Theorie noch das Stichprobendatenmaterial a l l e i n die Grundlage der Entscheidung für einen bestimmten Kurventyp bilden kann, sondern daß ein Kompromiß zwischen beiden gefunden werden muß.[1] Die ökonomische Theorie kann häufig Angaben darüber machen, welche Variablen relevant sind und wie die Funktionsform u n g e f ä h r aussehen muß, selten aber führt sie zu einer ganz konkreten Form. Häufig soll es sogar die Aufgabe der statistischen Untersuchung sein, anhand empirischen Zahlenmaterials zwischen zwei konkurrierenden Theorien zu entscheiden.[2] Das beobachtete Datenmaterial sollte niemals allein die Grundlage für die Wahl der Funktion bilden, da es - wie GOLDBERGER es formuliert - "niemals wirklich für sich selbst spricht".[3] Auch anhand unseres Rechenbeispiels im Abschnitt 6.3.2 (Tabelle 6.2), das im Abschnitt 6.3.3.3.2 noch einmal zur Prüfung der Linearität herangezogen wurde, haben wir gesehen, daß zumindest d r e i alternative Funktionsformen (Gerade, Parabel, Hyperbel) für d i e v o r l i e g e n d e n D a t e n zu vernünftigen Schätzwerten führten. Darum sollte es auch nicht das Ziel einer Regressionsanalyse sein, zufällig beobachtetes Datenmaterial möglichst perfekt durch die Regressionsfunktion wiederzugeben oder zu erklären, sondern es sollte vielmehr versucht werden, aufbauend sowohl auf theoretischen Kenntnissen und früheren Untersuchungen als auch auf dem zu analysierenden Zahlenmaterial

---

1) A.S. GOLDBERGER, a.a.O., S. 125 ff.

2) Ebenda. S. 126. Vgl. auch das Kapitel 3 dieser Arbeit.

3) A.S. GOLDBERGER, a.a.O., S. 126. (Übersetzung des Verfassers.)

einer Stichprobe,[1] eine Relation zu finden, die betriebswirtschaftlichen Gesichtspunkten Rechnung trägt und die die gesamte, und nicht nur eine für einen Ausschnitt gültige Beziehung zwischen den untersuchten Größen widerspiegelt.[2]

---

1) Theoretische Kenntnisse und empirisches Datenmaterial können als S u b s t i t u t e angesehen werden, d. h. je genauer (ungenauer) die Angaben sind, die die Theorie oder frühere Untersuchungen liefern, um so geringer (größer) wird man im allgemeinen den Stichprobenumfang wählen können (müssen).

2) Vgl. auch C.H. GOULDEN, a.a.O., S. 126.

7. Kapitel

EINIGE BESONDERE PROBLEME DER KOSTENUNTERSUCHUNG MIT HILFE DER REGRESSIONSANALYSE

7.1 Autokorrelation bei Zeitreihen

7.1.1 Begriff der Zeitreihe

Da bei Kostenuntersuchungen das Datenmaterial häufig in Form von Zeitreihen vorliegt, soll in diesem Abschnitt auf die damit verbundenen Probleme eingegangen werden.

Unter einer Zeitreihe versteht man einen Satz von Beobachtungen einer örtlichen Einheit zu verschiedenen aufeinander folgenden Zeiträumen oder Zeitpunkten, wobei jede Beobachtung in einer Quantität und in einer Angabe über die Zeit, in der diese Quantität vorlag, besteht.[1] So sind zum Beispiel die aufgezeichneten Beobachtungsreihen bei der Verteilung der Stromkosten auf die Kostenträger[2] Zeitreihen, ebenso die von MATT verwendeten Daten für die Analyse des Dampfverbrauchs[3] sowie die im Rahmen der Sortimentsplanung verwendeten Zahlen[4] und ähnliche Beispiele mehr, die im Verlauf dieser Arbeit untersucht wurden. Lediglich bei der

1) Vgl. E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 205.; W.A. WALLIS u. H.V. ROBERTS, a.a.O., S. 465.

2) Vgl. die Tabelle auf S. 47.

3) Vgl. S. 107 ff.

4) Vgl. S. 126 ff.

Analyse der Stückkosten im zwischenbetrieblichen Kostenvergleich lag das Datenmaterial in Form von sog. "örtlichen" statistischen Reihen vor,[1] da hier die Korrelation von Kosten und Einflußgrößen für verschiedene Betriebe während des g l e i c h e n Zeitraums erfolgte.

Obwohl sich Korrelationen von Zeitreihen rein rechnerisch ebenso wie andere Korrelationen durchführen lassen, treten hier Besonderheiten auf, die im folgenden kurz dargestellt werden sollen.

7.1.2 Probleme bei Zeitreihen

Da die Glieder einer Zeitreihe in einer zeitlichen Verbundenheit miteinander stehen, ist häufig eine Grundtendenz der Entwicklung, ein T r e n d festzustellen,[2] der bedingt, daß die Reihenglieder nicht im wahrscheinlichkeitstheoretischen Sinn voneinander unabhängig sind.[3] Diese Verbundenheit oder Korrelation der Reihenglieder untereinander bezeichnet man als "A u t o k o r r e l a t i o n" oder ausführlicher als "Autokorrelation der Residuen", da die Verteilung der Residuen um die Regressionsgerade bei Vorliegen von Autokorrelation nicht mehr zufällig ist.[4]

In diesem Fall ist insbesondere bei Schätzungen mit den nach der Methode der kleinsten Quadrate ermittelten Regressionskoeffizienten Vorsicht geboten, da bei starker Autokorrelation in den zu korrelierenden Reihen "die immer wie-

1) Vgl. Tabelle 6.5 auf S. 193.
2) E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 205.
3) Vgl. G. TINTNER, a.a.O., S. 288.
4) Ebenda. S. 48.

derkehrenden Beziehungen an den einzelnen Punkten der beiden statistischen Reihen .. immer wieder auftretende Bestätigungen für eine vorliegende Korrelation zu sein [scheinen]";[1] obwohl aber in Wirklichkeit ein und derselbe Fall mehrmals beobachtet worden sein kann und es auf Grund der Autokorrelation nur so erscheint, als ob mehrere Fälle derselben Erscheinung vorlägen.[2]

7.1.3 Möglichkeiten zur Feststellung, Verringerung oder Beseitigung der Autokorrelation

Aus den genannten Gründen sollten Zeitreihen, bei denen das Auftreten von Autokorrelation vermutet wird, stets auf die Stärke dieser Autokorrelation untersucht werden. Das kann mit Hilfe von "Autokorrelationskoeffizienten" - zyklisch oder nichtzyklisch - geschehen, indem die Werte einer statistischen Reihe mit den um einen "Lag" verschobenen Werten derselben Reihe korreliert werden.[3] So wird zum Beispiel der nichtzyklische Autokorrelationskoeffizient 1. Ordnung,[4] der die Autokorrelation zwischen den Reihen $x_1, x_2, x_3 \dots x_{n-1}$ und $x_2, x_3 \dots x_n$ mißt, nach folgender Formel ermittelt:

$$K_1 = \frac{\sum_{i=1}^{n-1} x_i x_{i+1} - \frac{\sum_{i=1}^{n-1} x_i \sum_{i=2}^{n} x_i}{n-1}}{\sqrt{\left(\sum_{i=1}^{n-1} x_i^2 - \frac{(\sum_{i=1}^{n-1} x_i)^2}{n-1}\right)\left(\sum_{i=2}^{n} x_i^2 - \frac{(\sum_{i=2}^{n} x_i)^2}{n-1}\right)}} \quad (7.1)$$

1) E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 207 f.

2) G. TINTNER, a.a.O., S. 289 f.; E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 208.

3) Vgl. W.A. WALLIS u. H.V. ROBERTS, a.a.O., S. 467 f.

4) 1. Ordnung, weil die ursprüngliche Reihe mit der um e i n e Zeiteinheit verschobenen Reihe korreliert wird.

Ebenso wie für andere Korrelationskoeffizienten lassen sich auch für Autokorrelationskoeffizienten Aussagen über die Zuverlässigkeit machen, mit der die Angaben über ihre Höhe erfolgen,[1] d.h. es kann unter der Annahme des Stichprobenmodells getestet werden, ob in der Grundgesamtheit Autokorrelation vorliegt oder nicht.

Während durch diese Autokorrelationskoeffizienten lediglich die Existenz und der Grad der Autokorrelation festgestellt werden kann, wurden auch verschiedene Verfahren entwickelt, um die Autokorrelation zu verringern oder auszuschalten. Hier ist unter anderem die "Differenzenmethode", auch als "Variate-Difference-Methode" bezeichnet, zu nennen, bei der anstelle der ursprünglichen Beobachtungen die ersten Differenzen verwendet werden, um die Autokorrelation auszuschalten.[2] Daneben sind die Methoden von WOLD und ORCUTT sowie autoregressive Transformationen anderer Art bekannt, auf die hier aber nicht weiter eingegangen werden soll.[3]

Eine weitere Möglichkeit besteht darin, die "Zeit" selbst als erklärende Variable in die Regressionsgleichung mit aufzunehmen, was unter Verwendung der Methode der kleinsten Quadrate zu genau demselben Ergebnis führt wie eine Untersuchung mit Hilfe der Abweichungen von berechneten Trends.[4]

---

1) Vgl. auch den Abschnitt 4.4.4 auf S. 69 ff.

2) Einzelheiten vgl. bei G. TINTNER, a.a.O., S. 49. u. S. 279.; H. v. SCHELLING: Die wirtschaftlichen Zeitreihen als Problem der Korrelationsrechnung. Mit besonderer Berücksichtigung der "lag correlation". (Heft 11 der Veröffentlichungen der Frankfurter Gesellschaft für Konjunkturforschung. Hrsg. E. Altschul). Bonn 1931. S. 47.; O. ANDERSON, a.a.O., S. 238.

3) G. TINTNER, a.a.O., S. 50 f., S. 294 u. S. 302 ff.

4) Ebenda. S. 49 u. S. 277.; G. GRÖNER, a.a.O., S. 67 ff.

Zwar sind die nach der Methode der kleinsten Quadrate bei Vorliegen von Autokorrelation ermittelten Regressionskoeffizienten weiterhin erwartungstreu, aber um Konfidenzintervalle zu berechnen oder um Hypothesen zu prüfen, müssen die ermittelten Standardfehler der Regressionsparameter mit einem die Autokorrelation berücksichtigenden Faktor korrigiert werden,[1] da sie bei Vorliegen von Autokorrelation erheblich "unterschätzt", d.h. kleiner errechnet werden können, als sie in Wirklichkeit sind. Nach Korrektur der Standardfehler läßt sich für Regressionsschätzungen näherungsweise die Studentverteilung anwenden.[2]

Glücklicherweise spielt die Autokorrelation bei den in dieser Arbeit aufgezeigten Kostenuntersuchungen im betrieblichen Rahmen nicht d i e Rolle wie zum Beispiel bei makroökonomischen Untersuchungen von Kosten-, Angebots- oder Nachfragefunktionen, bei denen die in die Analyse eingehenden Daten durch allgemeine wirtschaftliche Aufschwungs- oder Depressionstendenzen viel stärker beeinflußt werden,[3] und bei denen auf Grund des oft langfristigen Charakters dieser Zeitreihen Trends, Konjunktur- und Saisonschwankungen von Bedeutung sind.[4] Trotz dieser Einschränkung bleibt auch für innerbetriebliche Kostenuntersuchungen die Gefahr des Auftretens von Autokorrelation bestehen.[5]

Bei den im Rahmen dieser Arbeit durchgeführten Kostenuntersuchungen kann man die Trendkomponente unberücksichtigt lassen, weil in aller Regel sehr kurzfristige Zeitreihen

---

1) G. TINTNER, a.a.O., S. 302.

2) Ebenda.

3) Vgl. auch G. GRÖNER, a.a.O., S. 6.

4) Ebenda. S. 49.; E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 206.

5) Vgl. S. 220 ff., wo auf Saisonschwankungen hingewiesen wird.

betrachtet werden, deren Zahlenmaterial noch dazu aus Grundgesamtheiten stammt, bei denen das Vorliegen eines Trends auch bei längerfristigen Analysen recht unwahrscheinlich ist.[1] Wir haben nämlich regelmäßig Produktionszahlen eines Betriebes, einer Kostenstelle, einer Maschine usw. zugrunde gelegt, die zwar im Laufe der Zeit schwanken - diese Schwankungen sind ja sogar der Grund für die Durchführung der Regressionsanalyse - bei denen aber auf Grund einer bestimmten Kapazitätsgrenze kein ausgeprägter Trend möglich ist. Wenn sich durch Kapazitätsausweitungen ein solcher Trend in Richtung auf immer höhere Produktionszahlen zeigen sollte, dann sind schon aus anderen Gründen - Vergleichbarkeit der Ergebnisse, Konstanz der Ausgangsdaten usw. - die bisher gültigen Regreßwerte mit Vorsicht zu verwenden. In diesen Fällen wird in der Regel eine neue Analyse, die auf den veränderten Gegebenheiten aufbaut, sinnvoll sein.

Dagegen kann auch im Rahmen von Kostenuntersuchungen eine Art von "Saisonkomponente" einen störenden Einfluß auf die Regressionsergebnisse ausüben. Dies kann bei allen Kosten (oder der entsprechenden Mengenkomponente) der Fall sein, die in irgendeiner Weise von der Jahreszeit abhängig sind, man denke wieder an die schon mehrfach als Beispiel verwendeten Energiekosten, wie Kohle-, Strom- oder Ölverbrauch für Heizungszwecke, Dampfverbrauch oder den Einsatz von Lösungs- oder Reinigungsmitteln. An anderer Stelle wurde schon beschrieben, wie mit Hilfe einer multiplen Regressionsanalyse unter Umständen der Einfluß der Jahreszeit - hier war es die Lufttemperatur - berücksichtigt werden kann.[2]

---

1) Vgl. auch die in der Praxis durchgeführten Analysen von FÖRSTER, der in einem Industriebetrieb und bei einer Baufirma die Veränderung der Kostenarten in Abhängigkeit vom Produktionsvolumen untersuchte und entweder gar keinen oder nur einen sehr geringen und deshalb zu vernachlässigenden Trend feststellte. E. FÖRSTER: Zur Anwendung ..., a.a.O., S. 563.

2) Vgl. S. 43.

Eine weitere Möglichkeit bietet sich durch die Einbeziehung einer "S c h e i n v a r i a b l e n" in die Regressionsgleichung.[1] So läßt sich zum Beispiel bei der Analyse von Elektrizitätskosten zum Zweck der Planung und Budgetierung die Jahreszeit durch Scheinvariable berücksichtigen, indem man für den Frühling und Herbst, für den Sommer und für den Winter je eine Variable in die multiple Regressionsgleichung aufnimmt.[2]

In einer Regressionsgleichung der Form

$$y' = a_{y.1234} + b_{y1.234}x_1 + b_{y2.134}x_2 + b_{y3.124}x_3 + b_{y4.123}x_4$$

würden die einzelnen Variablen dann folgende Bedeutung haben:

y' Schätzwert für den Stromverbrauch in kwh,

x_1 z.B. Maschinenlaufzeit in Stunden,

x_2 1. Scheinvariable, es gilt:
- $x_2 = 1$ für Frühling oder Herbst,
- $x_2 = 0$ für Sommer und Winter,

x_3 2. Scheinvariable, es gilt:
- $x_3 = 1$ für Sommer,
- $x_3 = 0$ für jede andere Jahreszeit,

x_4 3. Scheinvariable, es gilt:
- $x_4 = 1$ für Winter,
- $x_4 = 0$ für jede andere Jahreszeit.

---

1) Vgl. auch A.S. GOLDBERGER, a.a.O., S. 112 ff.

2) Vgl. auch G.J. BENSTON, a.a.O., S. 667 u. S. 670.

Selbstverständlich wird eine Regressionsanalyse, in die exakte Werte für Lufttemperatur, Niederschlagsmenge oder ähnliche Witterungseinflüsse eingehen, genauere Schätzwerte liefern können als eine multiple Regressionsgleichung mit einer "echten" unabhängigen Veränderlichen und drei Scheinvariablen. Da aber in der Regel Aufzeichnungen hierüber nicht in ausreichendem Maß zur Verfügung stehen werden, wird man sich häufig mit dieser Form der Analyse behelfen müssen, wobei allerdings im konkreten Fall zu prüfen ist, ob die so ermittelten Regreßwerte zu genügend exakten, d.h. für betriebliche Dispositionszwecke ausreichend genauen Schätzwerten führen.

Wenn auch die Probleme, die die Analyse von Zeitreihen für die Regressions- und Korrelationsrechnung mit sich bringt, bei kurzfristigen innerbetrieblichen Kostenuntersuchungen in der Regel nicht so schwerwiegend sind wie bei langfristigen makroökonomischen Analysen, so sollten sie doch auch in diesem Rahmen Beachtung finden, um bessere Schätzwerte zu erhalten und um Fehlinterpretationen zu vermeiden.[1] Keinesfalls jedoch sollte man eine Anwendung der Regressions- und Korrelationsanalyse auf Zeitreihenmaterial grundsätzlich ablehnen, da sie bei Berücksichtigung des Charakters des Datenmaterials und eventuellen Korrekturen vor oder während der Analyse durchaus zu brauchbaren Ergebnissen führen kann.[2]

---

1) Vgl. auch den grundlegenden Artikel zur Problematik von Zeitreihenanalysen: G.U. YULE: Why Do We Sometimes Get Nonsense-Correlations between Time-Series? - A Study in Sampling and the Nature of Time-Series. "Journal of the Royal Statistical Society". 89 (1926) 1, S. 1 ff.

2) Vgl. E. FÖRSTER: Zur Anwendung..., a.a.O., S. 563.

7.2 Heteroskedasie

Eine weitere Voraussetzung für die Anwendung von Schätzverfahren nach der Methode der kleinsten Quadrate ist das Vorliegen von Homoskedasie. Eine Verteilung wird dann als homoskedatisch[1] bezeichnet, wenn die Varianzen der "Störungskomponenten"[2] oder Residuen, d.h. der durch die Regressionsfunktion nicht erklärbaren Veränderungen in der abhängigen Variablen, in allen Beobachtungsperioden gleich groß sind.[3] Anders ausgedrückt: Die Varianzen dürfen nicht von der Größe der y-Werte und der x-Werte abhängig sein. Im Fall nichtkonstanter Varianzen - als Heteroskedasie bezeichnet - können keine zuverlässigen Konfidenzintervalle für die Regressionskoeffizienten ermittelt werden,[4] da ihre Standardfehler verzerrt sind,[5] d.h. die Regressionsfunktion läßt sich nicht für Schätzzwecke unter Angabe einer statistischen Wahrscheinlichkeit verwenden.

Da Heteroskedasie auch bei regressionsanalytischen Kostenuntersuchungen auftreten kann,[6] - zum Beispiel bei zwischenbetrieblichen Kostenvergleichen, wenn die untersuchten Betriebe oder Kostenstellen in ihrer Größe wesentlich voneinander abweichen, so daß in Extremfällen die K o s t e n v a r i a t i o n e n einer Stelle größer sein können, als die gesamten Kosten einer anderen[7] - sollte das Datenmaterial untersucht werden und nicht einfach Homoskedasie von vornherein unterstellt werden.

---

1) auch als "homoskedastisch" bezeichnet.

2) In der anglo-amerikanischen Literatur als "disturbances" bezeichnet.

3) Vgl. R. E. JENSEN, a.a.O. S. 269.; G.J. BENSTON, a.a.O., S. 668.

4) Hier ist nur die Anwendung verteilungsfreier Verfahren möglich. Vgl. dazu S. 144 ff.

5) Vgl. G. J. BENSTON, a.a.O., S. 668.

6) Vgl. auch F.R. OLIVER, a.a.O., S. 69.

7) Vgl. das Beispiel bei R.E. JENSEN, a.a.O., S. 270.

Als relativ einfaches Verfahren wird von JENSEN der Test von BARTLETT für Varianzenhomogenität vorgeschlagen, der auf elektronischen Datenverarbeitungsanlagen mit geringem Zeitaufwand durchzuführen ist.[1)]

Da die mit dem Auftreten von Heteroskedasie verbundenen Probleme noch nicht vollständig gelöst sind, ist man weitgehend auf Transformationen des Datenmaterials angewiesen, um zu konstanten Varianzen zu gelangen.[2)] Hierbei kann gerade bei ökonomischen Daten die Transformation der Variablenwerte in logarithmische Größen von Nutzen sein,[3)] wobei die Wirksamkeit dieses Verfahrens mit Hilfe eines Schaubildes, in dem die Werte der unabhängigen Variablen und die Residuen eingetragen werden, getestet werden kann.[4)]

Da schon das getrennte Auftreten von Multikollinearität,[5)] von Autokorrelation oder Heteroskedasie schwierige Probleme für eine sinnvolle und aussagefähige Regressionsanalyse mit sich bringt, läßt ihre gemeinsame Existenz im Ausgangsdatenmaterial eine Ausschaltung der dadurch entstehenden Verzerrungen fast unmöglich erscheinen, so daß die auf der Basis eines solchen Datenmaterials gewonnenen Ergebnisse nur geringen Wert besitzen. Es muß deshalb noch einmal betont werden, daß es wichtig ist, die Beobachtungswerte v o r Durchführung der eigentlichen Rechnung daraufhin zu überprüfen, ob die Modellvoraussetzungen der Regressionsanalyse wenigstens näherungsweise als gegeben angenommen werden können.

---

1) Vgl. das Beispiel bei R.E. JENSEN, a.a.O., S. 270 und die dort angegebene Literatur.

2) Ebenda. S. 270 f.; G.J. BENSTON, a.a.O., S. 668.

3) Vgl. z.B. F.E. CROXTON u. D.J. COWDEN, a.a.O. S. 613 ff.; M. EZEKIEL u. K.A. FOX, a.a.O., S. 73 u. S. 93 ff., und den Anhang S. 252 ff.

4) G.J. BENSTON, a.a.O., S. 668.

5) Vgl. dazu den Abschnitt 5.3.4., S. 111 ff.

Die Gefahr für den Kostenrechner liegt bei Anwendung der Methode der kleinsten Quadrate - wie bei den meisten statistischen Verfahren - in der Tatsache, daß die Rechenoperationen in gleicher Weise durchgeführt werden können, gleichgültig, ob alle Voraussetzungen für eine s i n n v o l l e Analyse gegeben sind oder nicht.[1] Jeder Kostenrechner, der die Regressions- oder Korrelationsanalyse für seine speziellen Rechnungszwecke einsetzen will, muß sich der Tatsache bewußt sein, daß die mit Hilfe dieser Verfahren ermittelten Ergebnisse nie besser sein können als das Datenmaterial, aus dem sie gewonnen wurden.

7.3 Gewinnung des Datenmaterials

7.3.1 Länge der Beobachtungszeiträume - sachliche und zeitliche Abgrenzung des Datenmaterials

Bei der Ermittlung des der Korrelations- und Regressionsanalyse zugrunde liegenden Datenmaterials sind zunächst Überlegungen bezüglich der Länge der Beobachtungszeiträume anzustellen. Es geht hier um die Frage, inwieweit sich die Erfassung von täglichen, wöchentlichen, monatlichen oder jährlichen Beobachtungsdaten auf die Ergebnisse der Analyse auswirkt.

Grundsätzlich läßt sich feststellen, daß keine allgemeingültige Aussage über eine optimale Länge der Erfassungszeiträume getroffen werden kann, daß vielmehr eine Konfliktsituation bei der Wahl zwischen kurzen und langen Beobachtungs-

1) Sieht man einmal von dem Fall ab, daß die Multikollinearität in "offener Form" auftritt. Auf S. 111 f. wurde dargestellt, warum das Gleichungssystem dann nicht lösbar ist.

perioden besteht. Ein Argument für die Wahl möglichst langer Zeiträume bildet die Tatsache, daß häufig "Lags" zwischen der buchmäßigen Aufzeichnung von Kosten und Produktions- oder Leistungseinheiten bestehen. Man denke zum Beispiel an die Berichterstattung von Löhnen, die in manchen Betrieben nur im Zwei-Wochen-Turnus oder monatlich aufgezeichnet werden, wenn nämlich die Lohnzahlung erfolgen soll. In diesem Fall kann man sich bei Durchführung einer Regressionsanalyse auch nur auf diese Monats- bzw. Zwei-Wochen-Zahlen beziehen, selbst wenn die erbrachte Leistung wöchentlich oder gar tageweise festgehalten wird. Der Lag zwischen der Berichterstattung von Lohnkosten und Produktionszahlen kann also nur durch die Wahl eines längeren Beobachtungszeitraums ausgeschaltet werden. Wählt man dagegen für die Analyse tägliche Werte, so müßten die Kosten als die abhängige Veränderliche auf tägliche Werte "umgerechnet" werden, was notwendigerweise mit mehr oder weniger genauen Verteilungsschlüsseln erfolgen müßte, wodurch das Datenmaterial von vornherein mit gewissen Ungenauigkeiten behaftet wäre.

Damit läßt sich als erstes Ergebnis festhalten, daß die Zeiträume, in denen die Erfassung des Datenmaterials erfolgt, l a n g g e n u g sein müssen, um solche Lags auszuschalten, die durch unterschiedliche Aufzeichnungs- oder Berichtszeitpunkte für abhängige oder unabhängige Variable entstehen.[1)]

Diesem Argument für die Wahl längerer Perioden steht die Tatsache entgegen, daß die Wahrscheinlichkeit für Veränderungen in der Höhe der u n a b h ä n g i g e n Variablen um so größer ist, je länger der Erfassungszeitraum gewählt

---

1) Vgl. auch G.J. BENSTON, a.a.O., S. 663.

wird. Da für diese Zeit jedoch nur e i n Wert für diese Variable in die Analyse eingehen kann, werden zwangsläufig die Veränderungen zu einem Durchschnittswert ausgeglichen. Auch hierdurch können Ungenauigkeiten in die Analyse hereingetragen werden, da möglicherweise durch die Durchschnittsbildung die wahre Beziehung zwischen Kosten und der oder den Einflußgrößen verschleiert wird.[1)]

Bei praktischen Untersuchungen im Rahmen der Kostenrechnung gilt es einen Kompromiß zu finden, der beiden Forderungen Rechnung trägt. Dabei wird es sich in manchen Fällen nicht vermeiden lassen, das Berichtswesen eines Betriebes von längerfristigen auf kürzerfristige - oder umgekehrt - Zeiträume umzustellen, um gleiche Bedingungen für die Beobachtungswerte von abhängiger und unabhängiger Veränderlicher zu schaffen.

In diesem Zusammenhang verdient ein weiterer Gesichtspunkt erwähnt zu werden. An mehreren Stellen dieser Arbeit - insbesondere bei der Ermittlung der Sollkostenkurve in der flexiblen Plankostenrechnung[2)] - wurde auf die Problematik der Verwendung von Vergangenheitszahlen für Gegenwarts- oder Zukunftsfragen eingegangen. Da ein Betrieb kein "stationäres Gebilde" ist, sondern im Zeitablauf stetigen Veränderungen unterworfen ist - sei es durch den Einsatz anderer Arbeitskräfte, durch die Verwendung neuer, modernerer Maschinen oder Materialien, durch die Erhöhung der Produktionskapazität, durch neue Fertigungsverfahren usw. - sind in der Regel die in der Vergangenheit beobachteten Werte nicht ohne weiteres für zukünftige Planungszwecke zu verwenden.[3)] Vielmehr ist es erforderlich, von Zeit zu Zeit

---

1) Vgl. auch G.J. BENSTON, a.a.O., S. 663.

2) Vgl. den Abschnitt 5.2.3.1 auf S. 86 ff.

3) Vgl. auch J.S. CHIU u. D.T. DE COSTER, a.a.O., S. 678.

das bisher benutzte Regressionsmodell zu überprüfen und - falls erforderlich - eine neue Analyse durchzuführen, die die veränderten Gegebenheiten berücksichtigt.[1)]

Zu der geforderten Konstanz des Datenmaterials gehören auch die betrieblichen Kostenrechnungsanweisungen. Es muß gewährleistet sein, daß keine grundlegenden Veränderungen zum Beispiel in den Aufzeichnungsmethoden für die Variablen, in der Kostenartengliederung, der Gemeinkostenverrechnung auf Kostenstellen und Kostenträger,[2)] der Periodenabgrenzung[3)] und der Abgrenzung der Zuständigkeits- oder Verantwortungsbereiche im Zeitablauf vorgenommen werden. Diese Kontinuität in Bezug auf die Entstehung des in die Analyse eingehenden Datenmaterials muß vor der Durchführung der Rechenoperationen geprüft werden, da - wie wir gesehen haben - die Beobachtungswerte eine Stichprobe aus einer Grundgesamtheit darstellen, die nur dann zu sinnvollen Ergebnissen führt, wenn alle Wertepaare - als die Elemente der Stichprobe - aus der gleichen Grundgesamtheit stammen.[4)] Diese Forderung setzt die Konstanz der oben genannten Bedingungen voraus.

Auch bei einer späteren Verwendung der Regressionsergebnisse ist die Einhaltung dieser Voraussetzungen zu überprüfen, weil zum Beispiel zwischenzeitliche Umstellungen auf andere

---

1) Vgl. E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 31.; E.E. COMISKEY, a.a.O., S. 238.

2) So dürfen z.B. keine Veränderungen in den Umlageschlüsseln oder Zuschlagsgrundlagen vorgenommen werden.

3) Man denke z.B. an die Verteilung von Urlaubslöhnen - monatlich mit einem Zwölftel des Jahresbetrages oder in Prozenten der monatlichen Bruttolohnsumme - für Meister und Vorarbeiter, die, da sie Gehaltsempfänger sind, häufig erst am Ende des Jahres einen Bonus oder eine Pauschalvergütung für geleistete Überstunden erhalten.

4) Vgl. auch P.R. MC CLENON, a.a.O., S. 544.

Kostenrechnungsverfahren - wie Direct Costing oder Deckungsbeitragsrechnung - Änderungen in den Abrechnungsverfahren bedingen, die die bisher gültigen Regreßwerte ihre Aussagekraft verlieren lassen. Ähnliches gilt für die Analyse von Kosten - im Gegensatz zu der Analyse von Verbrauchsmengen - wenn Preisänderungen eingetreten sind und die Preise der Kostengüter nicht als selbständige Variable in die Regressionsgleichung aufgenommen worden sind, sondern sich nur in den Kosten selbst niedergeschlagen haben.

Auch eventuelle Saison- oder Jahreszeiteinflüsse können die Bestimmung der Länge der Beobachtungsperioden beeinflussen. Neben der bereits erwähnten Einbeziehung von Scheinvariablen in die multiple Regressionsgleichung[1] lassen sich unter Umständen Jahreszeiteinflüsse auch durch eine sinnvolle Wahl der Beobachtungsperiode ausschalten. Bei Kostenarten, deren Höhe von der Jahreszeit mit beeinflußt wird, zum Beispiel Licht- und Heizungskosten, wählt man für die Regressionsanalyse einen Zeitraum von mindestens einem Jahr, wobei die einzelnen Beobachtungswerte sich auf Monatszahlen beziehen können. Auf diese Weise erreicht man, daß für spätere Planungs- oder Kontrollzwecke Durchschnittswerte vorhanden sind, die das Gesamtbild der Abhängigkeit der Kosten von ihren Einflußgrößen wiedergeben, ohne durch jahreszeitlich bedingte Temperaturunterschiede oder Lichtverhältnisse verzerrt zu sein,[2] da sich diese Unterschiede im Laufe des Jahres von selbst ausgleichen.

7.3.2 Stichprobenumfang

Als weiterer Punkt verdient die Wahl des Stichprobenumfangs Beachtung. Da man bei Kostenuntersuchungen vielfach auf

---

1) Vgl. S. 221 f.

2) Vgl. auch W. KILGER, a.a.O., S. 359 f.

Zeitreihenanalysen angewiesen ist, ist der Stichprobenumfang identisch mit der Anzahl der in die Analyse aufgenommenen Beobachtungsperioden.[1] Rein rechnerisch ist eine Regressionsanalyse nach der Methode der kleinsten Quadrate durchführbar, wenn e i n Freiheitsgrad vorhanden ist, d.h. wenn der Stichprobenumfang um eins größer ist als die Zahl der unabhängigen Variablen.[2] Allerdings ist die Durchführung einer solchen Analyse in der Praxis wertlos, da die Ergebnisse auf Grund des dann auftretenden breiten Konfidenzintervalls für die Koeffizienten und Regreßwerte für Schätzungszwecke nicht zu verwenden sind, wenn man einen üblicherweise geforderten statistischen Sicherheitsgrad von mindestens 95 % voraussetzt. Um zu sinnvollen, in der Praxis verwertbaren Ergebnissen zu gelangen, muß also ein Stichprobenumfang gewählt werden, der die Anzahl der in die Analyse aufgenommenen unabhängigen Variablen weit übertrifft.[3] In diesem Fall verringert sich die Breite der Konfidenzintervalle - gleiches Signifikanzniveau vorausgesetzt - da die Standardfehler - als die eine Komponente für die Bestimmung der Konfidenzintervalle[4] - etwa um die Quadratwurzel aus der Zahl der zusätzlich in die Analyse aufgenommenen Beobachtungsperioden reduziert werden.

In der Praxis können sich insoweit Schwierigkeiten ergeben, als nicht immer genügend viele Beobachtungswertepaare vorliegen werden. Häufig existieren keine oder nur wenige verwertbare Aufzeichnungen, insbesondere über die Höhe der un-

---

1) Bei der Analyse von "örtlichen" statistischen Reihen ist der Stichprobenumfang gleich der Anzahl der örtlichen Einheiten in einem Zeitpunkt oder Zeitraum. Vgl. auch S. 215 f.

2) Vgl. auch J.S. CHIU u. D.T. COSTER, a.a.O., S. 679.; G.J. BENSTON, a.a.O., S. 663 u. S. 154 dieser Arbeit.

3) Vgl. auch E.E. COMISKEY, a.a.O., S. 235.

4) Die andere Komponente ist der z-Wert für den Sicherheitsgrad der Aussage, der aus der Tabelle der Normalverteilung, der Studentverteilung oder ähnlichen Tabellen entnommen wird. Vgl. Abschnitt 5.6. S. 141 ff.

abhängigen Variablen. Hierin besteht der wesentliche Nachteil von ökonomischem Datenmaterial gegenüber naturwissenschaftlichen Daten, die in der Regel durch Experimente in großer Anzahl gewonnen werden können, wobei außerdem störende Einflüsse von anderen als den untersuchten Größen weitgehend ausgeschaltet werden können.[1] In der ökonomischen Praxis dagegen muß man sich mit dem Zahlenmaterial begnügen, das in tatsächlichen betrieblichen Situationen entstanden ist, ohne es durch Experimente erweitern zu können. Dadurch kommt einem guten, d.h. umfassenden und differenzierten betrieblichen Rechnungswesen eine besondere Bedeutung bei der Bereitstellung der Unterlagen für die Regressionsanalyse zu.

7.3.3 Streuungsintervall der unabhängigen Veränderlichen

An verschiedenen Stellen dieser Arbeit wurde auf das Problem der Extrapolation der durch die Regressionsanalyse gefundenen Beziehung zwischen den Kosten und ihren Einflußgrößen eingegangen, d.h. auf die Ausweitung der ermittelten Relation auf theoretisch denkbare oder zukünftig tatsächlich beobachtete Werte der unabhängigen Veränderlichen, die jenseits - sei es nach oben oder nach unten - der in die Analyse einbezogenen Werte liegen.[2] Besonders in Richtung auf niedrige, nahe an null liegende Werte der unabhängigen Variablen werden gern Extrapolationen oder Verlängerungen der Regressionslinien vorgenommen, um die Regressionskonstante "a" als fixe Kosten oder fixen Verbrauch interpretieren zu können. Im Verlauf dieser Arbeit wurde verschie-

1) Vgl. M. EZEKIEL u. K.A. FOX, a.a.O., S. 30 f.; R.G.D. ALLEN, a.a.O., S. 167.

2) Vgl. insbesondere S. 89 ff. u. S. 201 ff.

dentlich auf die Problematik dieser Vorgehensweise eingegangen, insbesondere bei der Verwendung von nichtlinearen Funktionsformen oder bei der Approximation solcher Funktionen durch eine Gerade. Es soll an dieser Stelle noch einmal betont werden, daß die nach der Methode der kleinsten Quadrate ermittelten Regressions- und Korrelationskoeffizienten und die Regreßwerte y' grundsätzlich nur für den Bereich Gültigkeit haben, den die beobachteten, in die Analyse aufgenommenen Werte für die unabhängige(n) Variable(n) umschließen. Aus dieser Tatsache folgt für die Auswahl des Datenmaterials die Forderung nach einem möglichst breiten Streuungsintervall für die unabhängige(n) Veränderliche(n). Wie wir gesehen haben, ist dieser Forderung bei praktischen Kostenuntersuchungen nur schwer Rechnung zu tragen, weil die x-Werte, die meistens in irgendeiner Form die Beschäftigung (Arbeitsstunden, Leistungseinheiten etc.) repräsentieren, meistens nur geringfügig um die normale Beschäftigungshöhe schwanken und daß Abszissenwerte, die nahe an dem Wert null liegen, fast nie beobachtet werden können. Sofern der Analysierende jedoch irgendeinen Einfluß auf das Datenmaterial nehmen kann, sollte er versuchen, für die Werte der unabhängigen Variablen einen möglichst großen Streubereich zu erhalten, damit bei späterer Verwendung der Regressionsergebnisse Extrapolationen weitgehend unnötig werden.[1)]

7.3.4 Meßfehler

Da das für die Regressionsanalyse der Kosten verwendete Datenmaterial aus tatsächlich in der Praxis vorgefundenen

1) Natürlich genügt es nicht, einen oder zwei außergewöhnlich hohe oder niedrige Werte der unabhängigen Variablen zu beobachten. Die x-Werte sollten vielmehr ziemlich gleichmäßig über den gesamten Streubereich verteilt sein. Sog. "Ausreißer" können das Ergebnis im Gegenteil sogar erheblich verfälschen.

Betriebssituationen stammt, sind Meßfehler bei der Aufzeichnung der beobachteten statistischen Reihen nicht auszuschließen.[1] Abgesehen von Fehlern, die durch falsches Ablesen von Zählern und Instrumenten, durch Übertragungsfehler bei der Übernahme in Berichtslisten, durch falsche Additionen und ähnliche Fehlerquellen entstehen, sind die Kosten und ihre Einflußgrößen häufig auch mit Ungenauigkeiten behaftet, die aus dem System der Kostenrechnung oder Betriebsabrechnung resultieren. Manche Kostenarten können in ihrer tatsächlichen Höhe gar nicht exakt erfaßt werden, sondern beruhen auf mehr oder weniger subjektiven Schätzungen, die in erhöhtem Maße Fehlern unterworfen sind.[2] Darüber hinaus können viele Kosten nicht exakt, d.h. verursachungsgemäß - wenn man die Gültigkeit des Kostenverursachungsprinzips unterstellt[3] - auf Kostenstellen oder Kostenträger verrechnet werden und müssen nach Umlage- oder Verteilungsschlüsseln den Verantwortungsbereichen oder Leistungseinheiten zugeteilt werden, wodurch häufig Verzerrungen in der "wahren" Kostenhöhe auftreten. Auch die Zurechnung von Kosten auf Abrechnungszeiträume bereitet teilweise Schwierigkeiten; man denke an die Abschreibungen für Wirtschaftsgüter des Anlagevermögens, die in der Regel zeitabhängig und nicht nach der Inanspruchnahme des Wirtschaftsgutes - also leistungsabhängig - verteilt werden. Ähnliche Probleme entstehen bei der Verrechnung von Leihkosten (Zinsen u.ä.), Steuern und Versicherungsprämien oder kalkulatorischen Wagniskosten.

Bei der Wertung dieser und ähnlicher Fehler in Bezug auf

---

1) Vgl. auch G.J. BENSTON, a.a.O., S. 664.

2) Man denke z.B. an kalkulatorische Wagniskosten, kalkulatorische Abschreibungen und Zinsen. Vgl. auch R.E. JENSEN, a.a.O., S. 273.

3) Vgl. dazu S. 60 ff. u. S. 77 ff.

die Aussagefähigkeit der Regressionsergebnisse muß man unterscheiden zwischen den Fehlerquellen und Ungenauigkeiten, die die Höhe der abhängigen Variablen, der Kosten, berühren und denjenigen, die Einfluß auf die unabhängige(n) Variable(n), die Kosteneinflußgrößen, nehmen. Beide Arten wirken sich auf die Verwendbarkeit der Regressionsergebnisse in unterschiedlicher Weise aus.[1)]

Grundsätzlich läßt sich feststellen, daß Fehler in der abhängigen Variablen nicht so schwerwiegend sind wie Fehler bei der Ermittlung der Werte für die die Kosten beeinflussenden Faktoren. Zwar kann hierdurch der Voraussagewert der Regressionsgleichung vermindert werden, die Schätzwerte für die Regressionskoeffizienten - seien es Stückkosten, Grenzkosten oder andere Größen - werden hierdurch jedoch nicht berührt,[2)] da durch Fehlerquellen dieser Art lediglich die "Störungskomponente" einer Veränderung unterworfen wird, die Regressionskonstante und der (die) Regressionskoeffizient(en) aber unverändert bleiben.[3)]

Größere Beachtung verdienen also Meßfehler, die die unabhängigen Veränderlichen betreffen, da in diesem Fall die Störungskomponente(n) mit der (den) unabhängigen Variablen korreliert ist (sind).[4)] Dadurch werden die durch die Stichprobe ermittelten Schätzwerte für die wahren Regressionskoeffizienten der Grundgesamtheit verzerrt, d.h. zu niedrig errechnet.[5)] Aus diesem Grund ist bei der Messung, der Aufzeichnung und eventuell erforderlicher Übertragung von Wer-

---

1) Vgl. auch G.J. BENSTON, a.a.O., S. 664 ff.

2) Ebenda. S. 665.

3) Ebenda. Fußnote 15, in der der Beweis für diese Aussagen angetreten wird.

4) Ebenda. S. 665. Fußnote 16.

5) Ebenda.

ten der unabhängigen Variablen besondere Sorgfalt erforderlich. In der Kostenrechnung gilt das insbesondere für die Durchführung von einfachen Regressionsanalysen, bei denen die unabhängige Variable durch die Produktion, die Beschäftigung, den Output - oder wie auch immer man diese Größe bezeichnet - repräsentiert wird. Diese Größe kann aber nur bei homogener Produktion[1] direkt ermittelt werden, während bei heterogener Leistungserstellung Umrechnungen erforderlich sind, um e i n e n Leistungs- oder Produktionsmaßstab zu erhalten. Ungenauigkeiten und Zurechnungsfehler bei der Ermittlung dieses "Generalproduktes" können die Ergebnisse der Regressionsanalyse erheblich verfälschen. Es zeigt sich an dieser Stelle erneut, daß die eigentliche Rechnung - zumal sie von programmierbaren Tischrechnern oder elektronischen Datenverarbeitungsanlagen in sehr kurzer Zeit durchgeführt wird [2] - nur einen relativ unbedeutenden Teil der gesamten Regressionsanalyse darstellt und daß Überlegungen und Untersuchungen v o r und n a c h Durchführung der Rechenoperationen eine wesentlich größere Bedeutung zukommt. Nur dann, wenn sich der Kostenrechner über die Beschaffenheit des Datenmaterials, das von Betrieb zu Betrieb anders geartet sein kann, Rechenschaft ablegt, kann er eine s i n n v o l l e Kostenuntersuchung mit Hilfe der Regressions- und Korrelationsanalyse durchführen.

---

1) Herstellung e i n e s Produktes.

2) Vgl. 9. Anhang I über die rechnerische Durchführung von Regressions- und Korrelationsanalysen, S. 252 ff.

8. Kapitel

ZUSAMMENFASSENDE INTERPRETATION DER UNTERSUCHUNGSERGEBNISSE IM HINBLICK AUF DIE ZIELSETZUNG

Der Verfasser hat sich bemüht, die Bedeutung der Regressions- und Korrelationsanalyse für die Lösung spezieller kostenrechnerischer Fragen darzustellen. Dabei wurde auch versucht, in den einzelnen Kapiteln auf die besonderen Vorteile, aber auch Probleme dieses Verfahrens einzugehen. Da eine anschauliche Darstellung nur mit Hilfe von zahlreichen Rechenbeispielen möglich war, mag in einigen Fällen die Systematik zugunsten der Anschaulichkeit etwas in den Hintergrund getreten sein. Aus diesem Grund erscheint es angebracht, abschließend eine systematische Zusammenfassung der bisherigen Teilergebnisse im Hinblick auf die Erreichung der Kostenrechnungsziele zu geben, wobei auch die Grenzen des gezeigten Verfahrens klar erkennbar werden sollen.

Im Abschnitt 1.2 wurde die Konsistenz der vorgegebenen Kostenrechnungsziele mit der Begründung nicht geprüft, daß diese Frage für die Beurteilung der Regressions- und Korrelationsanalyse im Zusammenhang mit kostenrechnerischen Problemen ohne Bedeutung ist. Die Rechtfertigung dieses Schrittes hat sich bereits aus den Teilergebnissen der vorhergehenden Kapitel ergeben. Bei allen Untersuchungen, gleichgültig ob sie sich auf die Ermittlung eines Perioden- oder Stückerfolgs, auf die Planung und Vorgabe oder auf die Kontrolle von Kosten bezogen, hat sich gezeigt, daß die Aufgabe der Regressions- und Korrelationsanalyse in der Formulierung und Darstellung von Abhängigkeiten oder Gesetzmäßigkeiten zwischen den Kosten bzw.

ihren Mengenkomponenten und deren Einflußgrößen liegt. Diese Darstellung, durch die erst eine Offenlegung der Kostenstruktur und damit des ursächlichen Zusammenhangs zwischen den Kosten und den sie beeinflussenden Variablen gelingt, ist notwendige Voraussetzung für die Erreichung s ä m t l i c h e r Kostenrechnungsziele. Insoweit kann man von einer universellen Anwendbarkeit der Regressions- und Korrelationsanalyse auf kostenrechnerische Probleme sprechen, so daß die Frage der Zielkonsistenz in der Kostenrechnung in diesem Rahmen nicht relevant wird.

Wenn dieser Arbeit trotzdem drei unterschiedliche Aufgabengebiete der Kostenrechnung vorangestellt wurden, so geschah das einmal, um einen Rahmen abzustecken, in dem die deskriptiven Untersuchungen durchgeführt werden konnten, zum anderen aber auch, weil die Regressions- und Korrelationsanalyse bei der Anwendung auf einzelne Ziele ganz besondere Merkmale - in positiver wie in negativer Hinsicht - aufweisen kann.

Wenn im folgenden die Ergebnisse anhand der im Abschnitt 1.2 genannten Kriterien "Objektivität, Flexibilität, Wirtschaftlichkeit, Zuverlässigkeit und Genauigkeit" zusammenfassend interpretiert werden, so haben diese Ausführungen grundsätzlich für alle drei untersuchten Hauptziele der Kostenrechnung Gültigkeit. Inwieweit bei einzelnen Aufgabengebieten unterschiedliche Aspekte auftreten, wird bei der Untersuchung der einzelnen Kriterien hervorgehoben.

Stellen wir den bedeutendsten Gesichtspunkt an die Spitze dieser Untersuchungen: Inwieweit gelingt durch die Anwendung der Regressions- und Korrelationsanalyse die Einschränkung von Subjektivismen und Ermessensspielräumen, d.h. eine Objektivierung in der Kostenrechnung? Die Unter-

suchungen haben gezeigt, daß die Durchführung der Regressions- und Korrelationsanalyse an die exakte Formulierung eines mathematischen Modells gebunden ist. Dadurch ist die Einführung beliebiger subjektiver Variationen der Modellprämissen ausgeschlossen. Das bedeutet, daß eine Untersuchung von jedem mit den rechnerischen Grundlagen dieses Verfahrens vertrauten Kostenrechner mit demselben Ergebnis nachvollziehbar ist. Nur eine derartige Methode bietet die Möglichkeit, subjektive Urteile weitgehend auszuschließen und Ermessensspielräume auf ein Mindestmaß einzuschränken. Dieser Gesichtspunkt erhält seine besondere Bedeutung aus der Tatsache, daß bei vielen konventionellen Verfahren der Kostenrechnung nicht unbedingt "Manipulationsmöglichkeiten" gesucht werden, sondern daß infolge Fehlens objektiver Entscheidungskriterien Ermessensspielräume z w a n g s l ä u f i g vorhanden sind. In diesem Zusammenhang sei erinnert an die Bestimmung der optimalen Schlüsselgröße für die Gemeinkostenverrechnung.[1] Es wurde gezeigt, daß in aller Regel eine Vielzahl von Bezugsgrößen für die Verrechnung der Gemeinkosten auf Kostenstellen und Kostenträger in Betracht kommt. Hier liefert die Regressions- und Korrelationsanalyse mit dem Korrelationskoeffizienten ein Maß, das die im Sinn des Auswahlkriteriums[2] beste Schlüsselgröße eindeutig determiniert, da seine Größe die Stärke des Einflusses widerspiegelt, den der untersuchte Faktor auf die Kosten ausübt. Damit bildet die Höhe des Korrelationskoeffizienten ein eindeutiges Entscheidungskriterium zwischen alternativen Bezugs-

---

1) Vgl. Abschnitt 4.4, S. 58 ff.

2) Vgl. hierzu insbes. S. 60 ff.

oder Schlüsselgrößen und ersetzt eine meist auf subjektiven Erwägungen beruhende Wahl nach "logisch erscheinenden" oder erfahrungsmäßig begründeten gedanklichen Kriterien. Die Aussagefähigkeit des Korrelationskoeffizienten erhöht sich durch die Einbeziehung von wahrscheinlichkeitstheoretischen Überlegungen, die es erlauben, Aussagen über Schätzintervalle und zugehörige Realisationswahrscheinlichkeiten zu gewinnen. Die Möglichkeit, für "zufällige" Ergebnisse bestimmte Wahrscheinlichkeiten angeben zu können, führt zu einer weiteren Einschränkung des Ermessensspielraums, da mathematisch-statistische Regeln den möglichen "Spielraum", d.h. Streuungsbereich, eindeutig festlegen.[1] Auf die besonderen Probleme, die bei der Gewinnung von Gemeinkostenschlüsseln mit Hilfe von Korrelationskoeffizienten auftreten, soll hier nicht näher eingegangen werden. Der an Einzelheiten interessierte Leser sei dazu auf den Abschnitt 4.4 verwiesen.

Die am Beispiel der Gemeinkostenschlüsselung dargestellte Möglichkeit, mit Hilfe von Regressions- und Korrelationsparametern zu einer größeren Objektivierung der Kostenrechnung zu gelangen, beschränkt sich natürlich nicht auf das genannte Beispiel. Die für die Durchführung dieses Verfahrens notwendige Gestaltung eines mathematischen Modells muß zwangsläufig auch bei der Anwendung auf Gebieten der Planung und Kontrolle von Kosten diesen Erfolg gewährleisten. Auch hier kann der an diesen Problemen besonders interessierte Leser auf die Untersuchungen der entsprechenden Abschnitte dieser Arbeit verwie-

---

1) Zum Problem der Bestimmung von Konfidenzintervallen und statistischen Realisationswahrscheinlichkeiten vgl. den Abschnitt 4.4.4, S. 69 ff. und insbes. den Abschnitt 5.6, S. 141 ff.

sen werden, insbesondere auf die Abschnitte 5.7 mit der Darstellung der Grundlagen für ein mathematisch formuliertes Betriebsmodell[1] und 6.2.2 ff., in denen die Möglichkeit zur Aufstellung einer Kostenkontrollkarte beschrieben wird.[2]

Nach Prüfung der Objektivität wollen wir uns der Frage der Flexibilität der Regressions- und Korrelationsanalyse und der mit ihrer Hilfe gewonnenen Ergebnisse zuwenden. Dabei soll zunächst untersucht werden, inwieweit das Verfahren selbst flexibel ist, d.h. inwieweit es sich auf die unterschiedlichen betrieblichen Situationen oder Fragestellungen anwenden läßt. Wir haben im 2. Kapitel, in dem die Grundlagen der Regressions- und Korrelationsanalyse dargestellt wurden, gezeigt, daß Regressionen in einfacher und multipler, in linearer und nichtlinearer Form durchgeführt werden können. Durch diese Möglichkeit der Abwandlung einer einfachen Linearregression kann die Regressionsanalyse gut an unterschiedliche oder sich im Zeitablauf ändernde Betriebssituationen oder Fragestellungen angepaßt werden. Wir wollen hier wieder ein Beispiel aus den bisherigen Untersuchungen herausgreifen, um diese Tatsache zu veranschaulichen. Bei der Durchführung eines zwischenbetrieblichen Kostenvergleichs wurde gezeigt, daß sich die untersuchte Abhängigkeit zwischen Stückkosten und Produktionsmengen durch eine lineare und zwei nichtlineare Funktionsformen darstellen ließ.[3] Wir kamen zu dem

---

1) Vgl. S. 161 ff.

2) Vgl. S. 174 ff.

3) Vgl. den Abschnitt 6.3, S. 190 ff.

Ergebnis, daß - je nachdem, welcher Zweck mit der Regressionsfunktion verfolgt wurde, z.B. möglichst gute Schätzwerte im beobachteten Bereich zu liefern oder den allgemeinen theoretischen Zusammenhang darzustellen - sowohl die Gerade als auch die Parabel als auch die Hyperbel als "beste" Lösung im Hinblick auf die verfolgte Zielsetzung angesehen werden konnten. Das Verfahren bietet demnach die Möglichkeit, beobachtetes Datenmaterial in unterschiedlicher Form darzustellen und damit verschiedenen Zielsetzungen Rechnung zu tragen. Trotz der erforderlichen Formulierung eines mathematischen Modells ist die Regressions- und Korrelationsanalyse also nicht an ein starres Schema gebunden, sondern weist über die Wahl alternativer Funktionsformen und die Einbeziehung mehrerer unabhängiger Variabler in multiplen Analysen eine relativ große Flexibilität auf.

Es bleibt zu fragen, wie flexibel die auf Grund bestimmten Datenmaterials gewonnenen Ergebnisse, d.h. die Regressions- und Korrelationsparameter und ihre Streuungsmaße, sind. Lassen sie sich auch in ähnlichen betrieblichen Situationen wie der untersuchten oder in folgenden Abrechnungsperioden verwenden oder ist bei Veränderungen stets eine neue Analyse erforderlich? Dieser Gesichtspunkt betrifft alle genannten Aufgabengebiete der Kostenrechnung, z.B. im Rahmen der Perioden- oder Stückerfolgsrechnung die Bestimmung von geeigneten Gemeinkostenschlüsseln[1] oder bei der Kostenkontrolle z.B. die Aufstellung von Kostenkontrollkarten[2]. Ganz besondere Bedeutung erhält dieses Problem jedoch bei der P l a n u n g von Kosten, die hier zukünftige Situationen durch aus Vergangenheitswerten gewonnene Regressions-

1) Vgl. S. 74 ff.
2) Vgl. S. 174 ff.

parameter wiedergegeben werden sollen. In diesem Zusammenhang ist insbesondere auf die Entwicklung der schrittweisen Regressionsanalyse hinzuweisen. Wir haben im Abschnitt 5.3 gesehen, wie durch die Berechnung alternativer Regressionen mit unterschiedlicher Zahl von unabhängigen Variablen die Berücksichtigung denkbarer zukünftiger betrieblicher Konstellationen in der Regel besser gelingt als bei den herkömmlichen Regressionsverfahren.[1)] Es wurde auch gezeigt, daß diese unterschiedlichen Regressionsgleichungen für verschiedene Betriebsabteilungen von besonderem Interesse sein können. Gerade die Möglichkeit, mit Hilfe der schrittweisen Regressionsanalyse nicht e i n e Regressionsgleichung zu berechnen, sondern je nach Anzahl der Schritte viele sog. "Zwischenregressionen" zu gewinnen, hat den Flexibilitätsgrad dieses Verfahrens wesentlich erhöht, zumal elektronische Datenverarbeitungsanlagen diese Gleichungen - bei möglicher Abwandlung der Regressionsform in Polynome verschiedener Grade - mit den entsprechenden Streuungsmaßen in kurzer Zeit bereitstellen. Auf diese Weise ist eine Art von "Schubladenplanung" möglich, d.h. die Erarbeitung einer Fülle von Konzepten, die auf unterschiedlichen zukünftigen Betriebssituationen aufgebaut sind und beim Eintritt einer derartigen Situation Verwendung finden können, ohne die Durchführung neuer Analysen erforderlich zu machen.

Aber auch die schrittweise Regressionsanalyse ist nicht in der Lage, sämtlichen zukünftigen Veränderungen im Betriebsablauf Rechnung zu tragen. Ständige umfangreiche Änderungen, die z.B. durch den Einsatz anderer Arbeitskräfte, durch die Verwendung modernerer Maschinen oder Materialien, durch die Erhöhung der Produktionskapazität, durch neue Fertigungsverfahren, aber auch durch die Ein-

---

1) Vgl. S. 99 ff., insbes. S. 105 ff.

führung anderer Kostenrechnungsmethoden bedingt sein können, machen eine dauernde Überprüfung des verwendeten Regressionsmodells notwendig und können von Zeit zu Zeit eine neue Analyse erfordern, die den geänderten Gegebenheiten Rechnung trägt.[1)]

Die Flexibilität der Regressionsergebnisse findet weiterhin in der Regel ihre Grenzen dort, wo versucht wird, Extrapolationen der Regressionsparameter vorzunehmen, d.h. die auf Grund eines bestimmten Datenmaterials gewonnenen Ergebnisse auf Bereiche auszuweiten, für die keine Angaben über die Variablen vorgelegen haben. In diesem Zusammenhang sei verwiesen auf die Interpretation der Regressionskonstanten[2)], auf die Darstellung der Gemeinkostenkurve bei Normal- und bei Überbeschäftigung[3)] und die besonderen Gefahren, die in einer Extrapolation der Regressionsergebnisse liegen, wenn versucht wird, eine Kurve für ein begrenztes Gebiet durch eine Gerade zu approximieren.[4)]

Bei Berücksichtigung der bisher genannten Vorteile der Regressions- und Korrelationsanalyse stellt sich die Frage, warum dieses seit langem bekannte Verfahren bisher noch nicht in größerem Maße Eingang in die Kostenrechnung gefunden hat. Es liegt die Vermutung nahe, daß - neben anderen Gesichtspunkten - insbesondere die vermeintlich hohen Kosten einer solchen Analyse für ihre relativ seltene Anwendung verantwortlich sind. Wir wollen im folgenden prüfen, ob die kostenrechnerische Anwendung von Regressions- und Korrelationsanalysen mit

---

1) Vgl. S. 225 ff.
2) Vgl. S. 50 ff., insbes. die Abb. 4.1 und 4.2.
3) Vgl. S. 89 ff., insbes. die Abb. 5.2.
4) Vgl. S. 201 ff.

einem wirtschaftlich vertretbaren Aufwand möglich ist. Dieses Problem betrifft zunächst einmal die Durchführung der Rechnung selbst, zum anderen aber auch die Bereitstellung des für die Analyse erforderlichen Datenmaterials.

Die bei Regressions- und Korrelationsanalysen mit einer großen Zahl von Beobachtungswerten erforderlichen umfangreichen Rechenoperationen können heute nicht mehr als Argument gegen die Anwendung dieses Verfahrens geltend gemacht werden. Durch die in den letzten Jahren ständig zunehmende Verbreitung von elektronischen Datenverarbeitungsanlagen entfallen zeitraubende und kostspielige manuelle Berechnungen. Von fast sämtlichen Herstellern von Datenverarbeitungsanlagen und auch von anderen Institutionen wird heute eine reichhaltige "software" an Regressionsprogrammen zur Verfügung gestellt, die es jedem Großbetrieb erlauben, diese Analysen durchzuführen.[1] Selbst kleineren Unternehmen ist mit Hilfe von programmierbaren Tischrechnern die wirtschaftliche Anwendung dieses Verfahrens möglich.[2] Wie im Anhang 9 gezeigt wird, wirken sich auch komplizierte Funktionsformen, die Ermittlung von Streuungsmaßen oder besonders umfangreiche Rechenoperationen, wie sie z.B. die schrittweise Regressionsanalyse erfordert, nicht negativ auf die Kosten der Rechnungsdurchführung selbst aus. Darüber hinaus erlauben die Regressions- und Korrelationsprogramme dem Kostenrechner eine Beschränkung auf die Analyse und Interpretation des Ausgangsdatenmaterials

---

1) Vgl. den Anhang 9, S. 252 ff., in dem die praktischen Durchführungsmöglichkeiten ausführlich dargestellt werden.

2) Vgl. S. 258 ff., wo diese Möglichkeiten aufgezeigt werden.

und der Ergebnisse, ohne die Rechnung selbst in allen Einzelheiten beherrschen zu müssen.

Vielschichtiger ist das Problem, inwieweit die durch die Bereitstellung des benötigten Datenmaterials entstehenden Kosten den wirtschaftlichen Einsatz der Regressions- und Korrelationsanalyse beeinflussen. Die Untersuchungen haben gezeigt, daß Aussagen über Regressions- und Korrelationsparameter mit einer um so größeren statistischen Wahrscheinlichkeit getroffen werden können, je größer die Zahl der in die Analyse aufgenommenen Wertepaare ist, d.h. je größer der Stichprobenumfang gewählt wird.[1] Bei Kostenuntersuchungen, denen in aller Regel Zeitreihen zugrundeliegen,[2] können sich Schwierigkeiten bei der Gewinnung großer Stichproben ergeben, da sich der Kostenrechner - im Gegensatz z.B. zum Naturwissenschaftler, der in der Regel die benötigten Daten in großer Zahl mit Hilfe von Experimenten beschaffen kann - auf das Zahlenmaterial beschränken muß, das in tatsächlichen betrieblichen Situationen entstanden ist. Wenn überhaupt eine Vergrößerung des Stichprobenumfangs möglich ist, so kann sie im allgemeinen nur durch Einbeziehung von weiter zurückliegenden Beobachtungswerten erreicht werden, wodurch der Aussagewert der Ergebnisse aber gerade für Planungszwecke beeinträchtigt werden kann.[3]

Neben diesen quantitativen Aspekt tritt das Problem der Qualität der Beobachtungswerte. Ungenaue oder mit Meßfehlern behaftete Daten können die Aussagefähigkeit der

---

1) Vgl. S. 74, S. 141 ff., insbes. aber S. 229 ff.

2) Vgl. den Abschnitt 7.1, S. 215 ff.

3) Vgl. auch S. 246.

Regressionsergebnisse einschränken. Hier können erforderlich werdende Neuaufzeichnungen oder Bereinigungen der Beobachtungsdaten zu erheblichen Kosten führen.

Die Gesichtspunkte ausreichender Quantität und Qualität des Ausgangsdatenmaterials haben grundsätzlich die gleiche Bedeutung für alle drei genannten Kostenrechnungsziele. Jedoch tritt insbesondere bei der Planung von Kosten mit Hilfe der Regressionsanalyse ein weiterer qualitativer Aspekt hinzu, der eine besondere Berücksichtigung der Wirtschaftlichkeitsfrage auf diesem Gebiet angebracht erscheinen läßt.

Wir wollen hierzu noch einmal die im Kapitel 5 untersuchten Alternativ- und Zielkalküle aufgreifen. Die in eine Regressionsanalyse eingehenden Wertepaare sind grundsätzlich in der V e r g a n g e n h e i t oder G e g e n w a r t beobachtete Realisationen der abhängigen und unabhängigen Variablen. Es erhebt sich die Frage, inwieweit diese Werte für Planungszwecke, d.h. für in die Z u k u n f t gerichtete Überlegungen verwendet werden können. Wir haben in den Abschnitten 5.3, in dem auf die Aufstellung von Kostenbudgets eingegangen wird (Zielkalküle), und 5.4, der die Gewinnung von Zahlen für kurzfristige Entscheidungen im Produktionsbereich behandelt (Alternativkalküle), gezeigt, daß hier durch die angewandten Verfahren - insbesondere durch die schrittweise Regressionsanalyse - in Verbindung mit der Überprüfung der Zuverlässigkeit der Schätzungen[1] in der Regel eine recht gute Übertragung der Vergangenheitswerte auf eventuell auftretende Änderungen in der Zukunft gelingt. Man kann in diesen Fällen meistens auf zusätzliche

---

1) Vgl. den Abschnitt 5.6, in dem ausführlich auf die möglichen Verfahren zur Einbeziehung von wahrscheinlichkeitstheoretischen Aspekten eingegangen wird.

Bereinigungen oder Aufbereitungen der Beobachtungswerte verzichten, so daß keine weiteren Kosten für die Analyse entstehen.[1)]

Ein weiterer Gesichtspunkt tritt bei der Gemeinkostenplanung, d.h. der Aufstellung von Sollgemeinkostenkurven im Rahmen der flexiblen Plankostenrechnung hinzu. Versucht man diese Kurven mit Hilfe von Regressionsfunktionen darzustellen, so weisen die in die Analyse aufgenommenen Vergangenheitszahlen in der Regel den Nachteil auf, daß in ihnen neben eventuellen anderen störenden Komponenten U n w i r t s c h a f t l i c h k e i t e n enthalten sind.[2)] Da aber der Sinn der Aufstellung von Gemeinkostenplänen gerade darin liegt, durch den Vergleich mit den tatsächlich entstandenen Istkosten Unwirtschaftlichkeiten in den Abteilungen oder Kostenstellen aufzudecken, tritt hier ein ernstes Problem für die regressionsanalytische Gemeinkostenplanung auf. Insbesondere KILGER und PACK haben auf diese Problematik aufmerksam gemacht und sehen darin das Hauptargument gegen die Konstruktion der Sollgemeinkostenkurve mit statistischen Mitteln[3)], weil nach ihrer Meinung nur durch eine gründliche und daher umfangreiche Aufbereitungsarbeiten umfassende Analyse die Eliminierung solcher Unwirtschaftlichkeiten gelingen kann. Die hierfür auftretenden Kosten können die Regressionsanalyse als wirtschaftlich nicht vertretbar erscheinen lassen, da sie in diesen Fällen

---

1) Es sei aber in diesem Zusammenhang noch einmal auf die Gefahren von Autokorrelation, Multikollinearität und Heteroskedasie hingewiesen. Vgl. S. 111 ff., S. 215 ff. und S. 223 ff.

2) Mit "Unwirtschaftlichkeiten" werden hier sämtliche Bestandteile der Istkosten bezeichnet, die über die als Verbrauchsnormen vorgegebenen Werte hinausgehen und vom Abteilungsleiter beeinflußt werden können.

3) Vgl. S. 86 ff. und W. KILGER, a.a.O., S. 372 ff. sowie L. PACK, a.a.O., S. 519.

ihren wesentlichen Vorteil gegenüber sog. "synthetischen" Verfahren der Gemeinkostenplanung, nämlich die kostengünstigere Ermittlung der Plankosten, verliert.[1)]

Bei der Beurteilung der Wirtschaftlichkeitsfrage sollte man nach Meinung des Verfassers jedoch nicht ausschließlich von theoretischen Forderungen an die Regressionsanalyse ausgehen, sondern gerade im Zusammenhang mit Gemeinkostenplanungen auch praktische Gesichtspunkte berücksichtigen. Die Forderung nach kostentheoretisch-gedanklich exakter Planung der Gemeinkosten und die Möglichkeit ihrer praktischen Realisation können nicht immer zu einem akzeptablen Kompromiß geführt werden. Es sei hierzu auf die statistischen Erhebungen in den U.S.A. verwiesen,[2)] die diesen Aspekt bestätigen. Hier hat sich nämlich gezeigt, daß regressionsanalytisch ermittelte Gemeinkosten wesentlich kostengünstiger bereitgestellt werden konnten als dies mit gedanklichen (synthetischen) Neuplanungen der Kosten möglich war, für die auch häufig weder Fachleute noch genügend Zeit zur Verfügung standen. Die Regressionsanalyse kann daher - auch in diesem Zusammenhang - wohl als wirtschaftlich vertretbares Hilfsmittel zur Erreichung akzeptabler Näherungslösungen angesehen werden.[3)]

Eng verbunden mit dem wirtschaftlichen Aspekt der Regressions- und Korrelationsanalyse ist die Frage der Zuverlässigkeit der durch sie gewonnenen Ergebnisse.

---

1) Zu den einzelnen Vorteilen synthetischer Gemeinkostenplanungen vgl. S. 83 und S. 88 f.

2) Vgl. S. 87 und S. 92 ff.

3) Auf den Grad der Genauigkeit dieser Schätzungen wird an späterer Stelle noch einzugehen sein.

Je mehr Aufbereitungsarbeiten am Ausgangsdatenmaterial vorgenommen werden, d.h. je mehr die Ausschaltung von störenden Komponenten wie z.B. Fehlkontierungen oder Meßfehlern bei der Aufzeichnung der Daten gelingt, um so zuverlässiger werden die Ergebnisse das tatsächliche Betriebsgeschehen widerspiegeln. Andererseits werden dadurch natürlich die Kosten für die Anwendung dieses Verfahrens steigen. Auch die Quantität des Zahlenmaterials, d.h. die Größe des Stichprobenumfangs, wirkt sich auf die Zuverlässigkeit regressionsanalytischer Schätzungen aus. Im Abschnitt 5.6 wurde gezeigt,[1] daß die Regreßwerte den Charakter von Erwartungswerten (Mittelwerten) haben, die in bestimmten Grenzen "streuen" können. Es wurde darauf hingewiesen, wie sich diese Streubereiche berechnen lassen und daß der Umfang der Stichprobe ein Faktor ist, der über die Streuungsmaße Einfluß auf die Größe dieser Konfidenzintervalle nimmt.[2]

Es gilt also in der Regel einen Kompromiß zu finden zwischen größerer Zuverlässigkeit der Regressionsergebnisse und höheren Untersuchungskosten, wobei allerdings die Regressions- und Korrelationsanalyse den Vorteil aufweist, mit Hilfe der Wahrscheinlichkeitsrechnung die Zuverlässigkeit der Schätzungen innerhalb bestimmter Bandbreiten zahlenmäßig angeben zu können.

Ein weiterer Aspekt soll lediglich am Rande vermerkt werden. Bedingt durch den bei praktischen Untersuchungen notwendigen Einsatz von programmierbaren Tischrechnern oder elektronischen Datenverarbeitungsanlagen sind Rechen-

---

1) Vgl. S. 141 ff.

2) Vgl. S. 154 und S. 230.

fehler bei der Durchführung der Analyse so gut wie ausgeschlossen, so daß die Ergebnisse vom rein rechnerischen Standpunkt aus unbedingt zuverlässig sind.

Die Verwendung von modernen Rechenanlagen führt auch zu einer hohen Genauigkeit bei der Berechnung der Regressions- und Korrelationsparameter. Die Koeffizienten und Streuungsmaße können mit einer praktisch beliebigen Anzahl von Dezimalstellen ermittelt werden. Darum ist es auch nicht dieses Problem, das die Frage der Genauigkeit der Regressionsergebnisse von Bedeutung erscheinen läßt. Vielmehr ist zu fragen, inwieweit man von Genauigkeit sprechen kann, wenn doch bei allen Größen ein mehr oder weniger großer Streubereich angegeben wird. Wir haben diese Frage bereits bei der Überprüfung der Zuverlässigkeit der Regressionsergebnisse gestreift.[1)] Hierzu ist noch einmal auf die Grundlagen eines Regressionsmodells bei Anwendung in der Kostenrechnung zu verweisen. Der Kostenrechner muß sich bei seinen Untersuchungen auf tatsächlich beobachtetes Datenmaterial beschränken und erhält damit lediglich eine S t i c h p r o b e aus einer Grundgesamtheit, d.h. aus allen möglichen Realisationen. Wir haben gezeigt, daß eine Stichprobe in der Regel keinen Wert liefern kann, der dem "wahren" Wert der Grundgesamtheit entspricht, sondern daß immer nur ein Konfidenzintervall angegeben werden kann, innerhalb dessen der Parameter der Grundgesamtheit mit einer bestimmten angebbaren Wahrscheinlichkeit zu finden ist. Dieses "Denken in Toleranzen" ist demnach keine Ungenauigkeit, die dem Verfahren anzulasten ist, sondern im Gegenteil die unter den in der Kostenrechnung vorgefundenen Gegebenheiten größtmögliche Präzisierung der berechneten Parameter. Gerade die Möglichkeit, die Streuungsbereiche und ihre Realisationswahrscheinlichkeiten eindeutig festzulegen, erhöht die Genauigkeit der Analyse.

---

1) Vgl. S. 249.

Man sollte sich jedoch stets Rechenschaft darüber ablegen, daß bis auf mehrere Stellen nach dem Komma "genau" ermittelte Ergebnisse wertlos sein können, wenn nicht mit der gleichen Genauigkeit bei der Gewinnung des in die Analyse aufgenommenen Datenmaterials vorgegangen wurde. Die Ergebnisse, die die Regressions- und Korrelationsanalyse liefert, können nie besser oder genauer sein als das Ausgangszahlenmaterial. Jedoch wird eine exakte Rechnung mit ungenauen Daten stets bessere Ergebnisse hervorbringen als subjektive Schätzungen oder gar Vermutungen, wenn man die Tatsache einer vorhandenen Ungenauigkeit bei der Interpretation der Ergebnisse berücksichtigt.

Wir haben gesehen, daß die Regressions- und Korrelationsanalyse ein vielseitig verwendbares Verfahren für den Kostenrechner ist, um Verrechnungen, Planungen und Kontrollen von Kosten durchzuführen, daß ihre sinnvolle Anwendung aber an bestimmte Voraussetzungen gebunden ist. Wenn diese "Spielregeln" eingehalten werden, kann dieses Verfahren zu wesentlichen Verbesserungen, insbesondere zu einer größeren Objektivität, in der Kostenrechnung beitragen.

9. Anhang I

RECHNERISCHE DURCHFÜHRUNG VON REGRESSIONS- UND KORRELATIONSANALYSEN

9.1 Manuelle Berechnungen

Während sich im Fall der linearen und auch der nichtlinearen Einfachregression die Regressionsparameter sowie der Korrelationskoeffizient und Streuungsmaße relativ leicht aus der Auflösung der Normalgleichungen und den für die anderen Maße angegebenen Formeln errechnen lassen, verursacht eine multiple Regressions- und Korrelationsanalyse - insbesondere bei Vorliegen umfangreichen Datenmaterials - schon einen erheblichen Rechenaufwand. So treten im Fall einer vierdimensionalen Regression (eine abhängige und drei unabhängige Variable) bereits vier Normalgleichungen auf, deren simultane Lösung von Hand zum Beispiel mit Hilfe der Determinantenrechnung neben dem Nachteil der mühsamen und zeitraubenden Errechnung der Koeffizienten noch das Problem aufweist, daß Rechenkontrollen während oder nach Beendigung der einzelnen Rechenschritte - um Rechenfehler oder -ungenauigkeiten sofort lokalisieren zu können - nicht möglich sind. Wenn nach Beendigung der gesamten Rechnung eine Kontrolle Fehler aufzeigt, sind s ä m t l i c h e Rechenoperationen zu wiederholen.

Aus diesen Gründen sind systematische Rechenschemata entwickelt worden, die eine relativ schnelle und kontrollierte Ausrechnung der gewünschten Regressions- und Korrelationsgrößen erlauben. Um dem Leser einen Einblick in den Aufbau solcher Rechenschemata zu verschaffen, sei hier kurz das

Rechenschema zur Ermittlung der partiellen Regressionskoeffizienten für eine vierdimensionale Regression mit Hilfe von GAUSSschen Multiplikatoren dargestellt.[1)]

Folgende Erläuterungen sind hierzu erforderlich:

Für die durch "S" gekennzeichneten Summen gelten folgende Beziehungen:

$$S_{11} = \sum_{i=1}^{n} (x_{i1} - \bar{x}_1) \cdot (x_{i1} - \bar{x}_1)$$

$$S_{1y} = \sum_{i=1}^{n} (x_{i1} - \bar{x}_1) \cdot (y_i - \bar{y})$$

oder in allgemeiner Schreibweise:

$$S_{jk} = \sum_{i=1}^{n} (x_{ij} - \bar{x}_j) \cdot (x_{ik} - \bar{x}_k)$$

für $j = 1, 2, \ldots, m$
und $k = 1, 2, \ldots, m$ und

$$S_{jy} = \sum_{i=1}^{n} (x_{ij} - \bar{x}_j) \cdot (y_i - \bar{y}).$$

Da $S_{jk} = S_{kj}$, brauchen im Rechenschema die Summen unterhalb der Diagonalen nicht eingetragen zu werden.

Die Werte der Spalten 1 bis 4 im auf den folgenden Seiten dargestellten Rechenschema, die in runde Klammern gesetzt sind, wurden aus vorhergehenden Zeilen übernommen und nicht nach den unter "Rechengang" angegebenen Berechnungsvorschriften ermittelt.

---

1) Entnommen aus E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 96. Hier findet man auch eine ausführliche mathematische Ableitung dieses Rechenschemas, vgl. S. 94 ff.

Rechenschema zur Ermittlung der partiellen Regressions-
koeffizienten mit GAUSS'schen Multiplikatoren[1)]

Zeile	Rechengang	1	2	3	4
01	Übernahme der Summen aus den Gleichungen	s_{11}	s_{12}	s_{13}	s_{1y}
02			s_{22}	s_{23}	s_{2y}
03				s_{33}	s_{3y}
04					0
10	Zeile (01) : s_{11}	1	b_{21}	b_{31}	b_{y1}
11	Zeile (02) - b_{21} · Zeile (01)	(b_{21})	s'_{22}	s'_{23}	s'_{2y}
12	Zeile (03) - b_{31} · Zeile (01)	(b_{31})		s'_{33}	s'_{3y}
13	Zeile (04) - b_{y1} · Zeile (01)	(b_{y1})			$-s^2_{1y} : s_{11}$

1) Vgl. E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 96.

Rechenschema zur Ermittlung der partiellen Regressions-
koeffizienten mit GAUSS'schen Multiplikatoren (1. Fortsetzung)

Zeile	Rechengang	1	2	3	4
20	Zeile (11) : S'_{22}	$-c'_{12}$	1	$-\frac{c_{23}}{c_{33}}$	$b_{y2.1}$
21	Zeile (12) - $(-\frac{c_{23}}{c_{33}})$ · Zeile (11)	$-\frac{c_{13}}{c_{33}}$	$(-\frac{c_{23}}{c_{33}})$	$\frac{1}{c_{33}}$	$\frac{b_{y3.12}}{c_{33}}$
22	Zeile (13) - $b_{y2.1}$ · Zeile (11)	$b_{y1.2}$	$(b_{y2.1})$		$-(b_{y1.2}\,S_{1y} + b_{y2.1}\,S_{2y})$
30	Zeile (21) : $\frac{1}{c_{33}}$	$-c_{13}$	$-c_{23}$	1	$b_{y3.12}$
31	Zeile (22) - $b_{y3.12}$ · Zeile (21)	$b_{y1.23}$	$b_{y2.13}$	$(b_{y3.12})$	$-(b_{y1.23}S_{1y} + b_{y2.13}\,S_{2y} + b_{y3.12}\,S_{3y})$

Die Regressionskonstante $a_{y.123}$ errechnet sich aus der Gleichung:[1]

$$a_{y.123} = \bar{y} - b_{y1.23}\bar{x}_1 - b_{y2.13}\bar{x}_2 - b_{y3.12}\bar{x}_3.$$

Eine weitere Methode wurde von M.H. DOOLITTLE entwickelt und ist unter demselben Namen als "DOOLITTLE-Methode" bekannt. Mit diesem Verfahren - auf das hier nicht weiter eingegangen werden kann[2] - lassen sich unter Verwendung von zahlreichen Rechenkontrollen sämtliche gewünschten Regressions- und Korrelationsmaße wie die partiellen Regressionskoeffizienten, Streuungsmaße wie der Standardfehler der Schätzung oder die Standardfehler der Regressionskoeffizienten für multiple Analysen ermitteln.

Auch diese Methoden erfordern bei Verwendung von nicht programmierbaren Tischrechenmaschinen noch einen erheblichen Rechenaufwand. Die in Abbildung 9.1 eingezeichneten Kurven geben einen Anhaltspunkt für die in etwa zu erwartenden Größenordnungen der erforderlichen Rechenzeiten.[3]

Die angegebenen Rechenzeiten gelten unter Verwendung einer Tischrechenmaschine mit Produktspeicherwerk für die Berechnung der aufgeführten Größen einschließlich des Zeitaufwands für Fehlersuche und Berichtigung.[4]

---

1) Diese Gleichung läßt sich aus der ersten Normalgleichung nach Division durch n und Auflösung nach $a_{y.123}$ gewinnen. Vgl. dazu auch den Abschnitt 2.2, S. 13 ff.

2) Für den an Einzelheiten dieser Methode interessierten Leser sei auf die Literatur hingewiesen. Ausführliche Darstellungen der DOOLITTLE-Methode findet man z.B. bei F.E. CROXTON u. D.J. COWDEN, a.a.O., S. 498 ff. mit einem Rechenbeispiel für eine Regressionsfunktion der Form $y' = a + bx + cx^2 + dx^3$. Außerdem bei M. EZEKIEL u. K.A. FOX, a.a.O., S. 489 ff. mit einem Rechenbeispiel für eine Regressionsfunktion der Form $y' = a + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5$.

3) Entnommen aus H. KNÜPPEL, A. STUMPF u. B. WIEZORKE: Mathematische Statistik in Eisenhüttenwerken. Teil I: Regressionsanalyse. "Archiv für das Eisenhüttenwesen". 29 (1958) 8, S. 521 ff., hier S. 526.

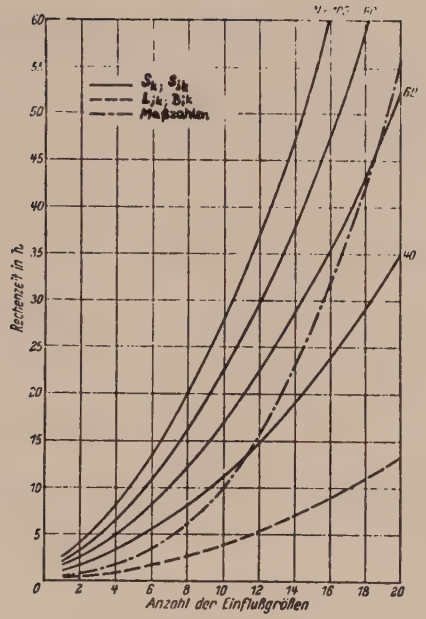


Abb. 9.1: Rechenzeiten zur Ermittlung der angegebenen Größen bei Einsatz einer Tischrechenmaschine. 1)

——— S_k, S_{jk} wobei gilt: $S_k = \sum x_k \quad k = 1, 2, \ldots, m.$

$$S_{jk} = \sum x_j x_k \qquad j,k = 1, 2, \ldots, m.$$

— — — L_{jk}, B_{jk} Die L_{jk} sind Gleichungskoeffizienten, die zur Berechnung der Regressionskoeffizienten benötigt werden.
Die B_{jk} sind die einfachen Bestimmtheitsmaße.

-.-.-.- Maßzahlen wie Regressionskoeffizienten usw.

Fortsetzung der Fußnote von Seite 256

4) Ebenda. Es sind laufende Rechenkontrollen eingebaut. Die Rechenarbeit ist so schematisiert, daß sie von Rechenkräften durchgeführt werden kann. Zur Angabe von Rechenzeiten vgl. auch J. KUCERA: Die Regressionsanalyse mit Hilfe kleiner Ziffernrechner. "Statistische Praxis. Zeitschrift für Statistik und Rechnungswesen". 18 (1963) 12, S. 354 ff.

1) Entnommen aus H. KNÜPPEL, A. STUMPF u. B. WIEZORKE, a.a.O., S. 526.

Für die Berechnung von S_k und S_{jk} ist durch Arbeitsteilung bei Einsatz mehrerer Rechenmaschinen eine Herabsetzung der Rechenzeit möglich. Für die Berechnung der L_{jk}, der B_{jk} und der Maßzahlen ist keine Arbeitsteilung mehr möglich.

9.2 Berechnungen mit programmierbaren Tischrechnern

Die außerordentlich hohen Rechenzeiten bei Verwendung einfacher Rechenmaschinen haben zu der Entwicklung von programmgesteuerten Berechnungsmethoden für die Regressions- und Korrelationsparameter geführt. Auf diese Weise lassen sich die Rechenzeiten von mehreren Stunden auf einige Minuten oder Sekunden reduzieren und zudem eine wesentlich größere Genauigkeit erzielen, als dies mit "normalen" Rechenmaschinen möglich ist.

Die für die Durchführung von Regressions- und Korrelationsanalysen von den Herstellern angebotenen Rechenanlagen reichen vom Tischrechner bis zu umfangreichen elektronischen Datenverarbeitungsanlagen, und die möglichen Rechenoperationen von der linearen Einfachregression bis zur schrittweisen Regressionsanalyse einschließlich Transformationen der unabhängigen Variablen.

Es ist bereits eine Reihe von programmierbaren Tischrechnern auf dem Markt. Da eine vollständige Aufzählung und Beschreibung ihrer Rechenfunktionen den Rahmen dieser Arbeit sprengen würde, soll stellvertretend für andere Maschinen kurz auf die von der Firma Olivetti entwickelte "olivetti programma 101" eingegangen werden.[1)]

1) Dieser Tischrechner wurde gewählt, da für ihn die ausführlichsten Beschreibungen statistischer Programme zur Verfügung standen: Statistische Programmsammlung für die "olivetti programma 101". Hrsg. von der Deutschen Olivetti GmbH., Direktion Marketing. Frankfurt a.M., Juli 1968.

Die "olivetti programma 101" arbeitet mit elektronischer Rechengeschwindigkeit nach Programmen, die auf Magnetkarten gespeichert werden können. Eine Programmiersprache im eigentlichen Sinn wird nicht verwendet: die Eingabe der Programmbefehle erfolgt über eine Funktionstastatur direkt in das "Gedächtnis" der Maschine.

Für die Durchführung der Regressions- und Korrelationsanalyse existieren im zentralen Programmarchiv der Deutschen Olivetti mehrere Programme. Der Bedienende hat also nur noch die Aufgabe, die Werte der einzelnen Variablen einzugeben, die Rechenoperationen für die Bestimmung aller gewünschten Größen werden von der Maschine anhand der Informationen aus dem gespeicherten Programm selbständig durchgeführt. Die eingegebenen Daten und die errechneten Resultate werden mit Dezimalpunkt und algebraischem Vorzeichen auf Papierstreifen ausgedruckt.

Nach den von der Deutschen Olivetti GmbH. zur Verfügung gestellten Unterlagen sind für die "olivetti programma 101" folgende Regressions- und Korrelationsprogramme vorhanden, wobei alle Schätzverfahren nach der Methode der kleinsten Quadrate durchgeführt werden:[1)]

1. Einfache lineare Regression der Form $y' = a + b\,x$ einschließlich der Berechnung des Produktmomentkorrelationskoeffizienten (BRAVAIS-PEARSON-Korrelationskoeffizient) und der Ermittlung einer t-verteilten Testgröße T für einen Signifikanztest des Korrelationskoeffizienten.[2)]
2. Nichtlineare Regression der Form $y' = a + b\,x + c\,x^2$ (Parabel).

---

1) Vgl. Statistische Programmsammlung für die "olivetti programma 101", a.a.O., S. 73 ff.

2) Die Gültigkeit der Nullhypothese wird dabei unterstellt. Vgl. auch S. 70.

3. Multiple Regression mit zwei unabhängigen Variablen (dreidimensionale Regression) einschließlich der Berechnung der einfachen und partiellen sowie des multiplen Korrelationskoeffizienten, wobei die Berechnung der partiellen und des multiplen Korrelationskoeffizienten aus den einfachen Koeffizienten erfolgt.[1)]
4. Logarithmische und halblogarithmische Regression und Korrelation für eine einfache Abhängigkeit.

9.3 Berechnungen mit elektronischen Datenverarbeitungsanlagen

Von vielen Herstellern größerer Datenverarbeitungsanlagen und vom Deutschen Rechenzentrum sind zahlreiche Programme zur Durchführung der Regressions- und Korrelationsanalyse entwickelt worden, die für die in der Praxis mit der Durchführung einer solchen Analyse Beschäftigten eine ganz erhebliche Vereinfachung bedeuten. Es ist nicht mehr erforderlich, ein eigenes Programm aufzubauen, da die auf dem Gebiet der Regressions- und Korrelationsrechnung bereitgehaltene Software so umfangreich und differenziert ist, daß nach Ansicht des Verfassers alle im Rahmen der Kostenrechnung auftretenden Regressionsprobleme mit diesen Programmen gelöst werden können.

Es kann an dieser Stelle natürlich keine lückenlose Aufzählung aller zur Zeit in Deutschland vorhandenen Regressions- und Korrelationsprogramme gegeben werden, da einerseits eine solche Übersicht zu umfangreich würde und zum andern laufend an der Entwicklung neuer Programme gearbeitet wird,

1) Vgl. zu den benutzten Formeln auch E. FÖRSTER u. F. EGERMAYER, a.a.O., S. 128 ff.

unter anderem auch von Industrieunternehmen, die eigene Programme für ausschließlich betriebsinternen Gebrauch erarbeiten.[1)]

Die folgenden Darstellungen sollen jedoch dazu dienen, einen Überblick über den Aufbau, die Arbeitsweise und die Ergebnisse sowie über die Bezugsquellen von bereits mehrfach benutzten Regressions- und Korrelationsprogrammen zu geben,[2)] die auch für die Lösung der Probleme des Kostenrechners wertvolle Dienste leisten können.

9.3.1 International Business Machines (IBM-System 360)

Für das IBM-System 360 existiert ein ganzes Programmpaket, das sog. "System/360 Scientific Subroutine Package (SSP)",[3)] eine Sammlung von über 200 Unterprogrammen zur Durchführung von zahlreichen Standardverfahren der Mathematik und Statistik. Sie sind in der Programmsprache FORTRAN IV geschrieben, jedoch anschließend im binären Maschinencode gespeichert, so daß eine Übersetzungszeit und damit auch Maschinenzeit gespart wird.[4)]

---

1) Diese Auskunft erhielt der Verfasser vom Vertriebsbüro Bielefeld der Firma "Remington Rand - Univac". Als Programmiersprache ist COBOL vorgesehen. Von den Firmen "Honeywell" und "NCR" wurde auf Anfragen nicht geantwortet, so daß deren Programme, die auch anderweitig nicht beschafft werden konnten, hier ohnehin fehlen.

2) Für Großbritannien wurde bereits im Jahr 1963 eine Aufstellung von multiplen Regressionsprogrammen für elektronische Datenverarbeitungsanlagen veröffentlicht. Vgl. dazu: M.J. GARSIDE: A Short List of Multiple Regression Programmes Available in Britain. "Applied Statistics. A Journal of the Royal Statistical Society". 12 (1963) 3, S. 151 ff.

3) IBM Application Program. System/360 Scientific Subroutine Package. (360 A - CM - 03 X). Version III. Programmer's Manual. Fourth edition. Hrsg. von International Business Machines Corporation 1966, 1967, 1968. Nr. H 20-0205-3.

4) Vgl. K. BONNET: Aufbau und Anwendungsmöglichkeiten des Programms "REGRE" (Regressions- und Korrelationsanalyse)

Aus diesen über 200 SSP-Unterprogrammen kann man durch ein Hauptprogramm mehrere Unterprogramme kombinieren und damit verschiedene statistische Probleme gleichzeitig lösen.

Eine solche Kombination stellt das Steuerprogramm "REGRE" dar, das zur Durchführung von multiplen linearen Regressionsanalysen geeignet ist. Es besteht aus den folgenden vier SSP-Unterprogrammen:[1)]

1. CORRE berechnet Mittelwerte, Standardabweichungen und eine Matrix von Produktmomentkorrelationskoeffizienten.[2)]

2. ORDER dient der Auswahl einer abhängigen Variablen und mehrerer unabhängiger Variabler aus einer größeren Anzahl von Variablen. Dazu wird eine Untermatrix von Interkorrelationen zwischen den unabhängigen Variablen gebildet.

3. MINV dient der Inversion der Korrelationsmatrix für die Variablen, die vom Unterprogramm ORDER ausgewählt wurden.

4. MULTR schließlich berechnet die Regressionskoeffizienten, die Korrelationskoeffizienten, Streuungsmaße usw.[3)]

Fortsetzung der Fußnote 4) von Seite 261:

4) aus der Programmbibliothek zum System IBM-360. Unveröff. Diplomarbeit am Institut für Ökonometrie und Wirtschaftsstatistik der Universität Münster. Münster 1969. S. 34.

1) IBM Application Program SSP, a.a.O., S. 17 u. S. 404 ff.

2) Ebenda. S. 36.

3) Ebenda. S. 37.

Neben diesen vier Unterprogrammen muß ein besonderes Unterprogramm DATA hinzugefügt werden, das zwar nicht im Steuerprogramm "REGRE" benötigt wird, aber von dem Unterprogramm CORRE aufgerufen wird.[1)]

Das Hauptprogramm "REGRE" gestattet Analysen mit bis zu 40 Variablen insgesamt, d.h. im Höchstfall mit 39 unabhängigen Einflußgrößen, und bis zu 99.999 Beobachtungswerten.[2)]

Eine Anlage vom Typ IBM 360/30 benötigt zur Durchführung der Analyse - wenn ein Kartenleser IBM 2540 und ein Drucker IBM 1403 Modell 3 verwendet werden - eine Zeit von 42 Sekunden. Bei Einsatz anderer Leser und Drucker kann sich die Zeitangabe entsprechend der Geschwindigkeit dieser peripheren Geräte verändern.

Im einzelnen werden die von "REGRE" ermittelten Regressions- und Korrelationsgrößen wie folgt ausgedruckt:[3)]

1. Mittelwerte für alle Variablen,
2. Standardabweichungen für alle Variablen,
3. Korrelationskoeffizienten zwischen den unabhängigen Variablen und der abhängigen Variablen[4)],

---

1) Vgl. K. BONNET, a.a.O., S. 35.

2) IBM Application Program SSP, a.a.O., S. 404. Dabei ist folgende Einschränkung zu machen: Es können bis zu 99.999 Beobachtungen aufgenommen werden, wenn die Beobachtungswerte durch das spezielle Eingabe-Unterprogramm DATA einzeln in die Maschine eingelesen werden. Wenn alle Daten vor der Berechnung der Korrelationskoeffizienten im Kernspeicher gesammelt werden sollen, hängt die Begrenzung der Anzahl der Beobachtungswerte von der Größe der Kernspeicherkapazität ab, die für die Eingabedaten vorhanden ist.(Sinngemäße Übersetzung des Verfassers).

3) Ebenda. S. 405.

4) Vgl. auch R. KNUSSMANN: ALGOL-Rechenprogramme statistischer Standardverfahren. "Unternehmensforschung - Operations Research". 5 (1961) 2, S. 61 ff., insbesondere S. 74 ff.

4. Regressionskoeffizienten,
5. Standardfehler der Regressionskoeffizienten,
6. t-Werte zum Testen der Regressionskoeffizienten,
7. Regressionskonstante,
8. Multipler Korrelationskoeffizient,
9. Standardfehler der Schätzung,
10. Tabelle zur Varianzanalyse für die multiple Regression,
11. Tabelle der Residuen, einschließlich der beobachteten y_i-Werte und der errechneten y'_i-Werte (wahlweise).

Neben diesem für l i n e a r e multiple Regressionsanalysen vorgesehenen Programm existiert ein weiteres Hauptprogramm, das die Durchführung nichtlinearer Analysen gestattet. Dieses als "POLRG" (Polynomial Regression) bezeichnete Programm setzt sich ebenfalls aus mehreren SSP-Unterprogrammen zusammen und erlaubt die Berechnung von Polynomen sukzessive ansteigender Grade.[1] Das Programm ist so eingerichtet, daß, wenn zwischen zwei aufeinander folgenden Graden von Polynomen keine Reduzierung der Summe der Abweichungsquadrate (Residuen) mehr eintritt, die Analyse automatisch beendet wird, ohne sie bis zum höchsten vorher festgelegten Grad durchzuführen.[2]

Auf die Darstellung der einzelnen SSP-Unterprogramme kann hier verzichtet werden, da sie prinzipiell nichts Neues gegenüber den Unterprogrammen für "REGRE" darstellen.

Die Kapazität des Programms "POLRG" erstreckt sich auf bis zu 100 Beobachtungen und bis zu Polynomen vom Grade 10. Bei Polynomen höherer Ordnung oder mehr als 100 Beobachtungswerten muß das vorliegende Programm modifiziert werden.

---

1) IBM Application Program SSP, a.a.O., S. 408.
2) Ebenda. (Sinngemäße Übersetzung des Verfassers).

Die Rechenzeit zur Durchführung der Analyse auf einer Anlage vom Typ IBM 360/30, dem Leser 2540 und dem Drucker 1403, Modell 3, wird mit 35 sec. angegeben.[1] Folgende Größen werden ausgedruckt:[2]

1. Die Regressionskoeffizienten für aufeinander folgende polynomische Grade, einschließlich der Regressionskonstanten,
2. Tabelle zur Durchführung einer Varianzanalyse für sukzessive polynomische Grade,
3. Tabelle der Residuen für die endgültige Regressionsfunktion (wahlweise) und
4. Tabelle der y_i-Werte und der y'_i-Werte (wahlweise).

Neben den Programmen "REGRE" und "POLRG" interessiert im Rahmen der Kostenrechnung insbesondere ein Programm zur Durchführung der schrittweisen Regressionsanalyse (stepwise regression). Ein solches Programm existiert unter dem Namen "STEPR", das die Auswahl der unabhängigen Variablen in der Reihenfolge ihrer Bedeutung für die Regressionsfunktion vornimmt.[3] Das Kriterium der "Bedeutung für die Regressionsgleichung" basiert auf der Reduzierung der Summe der Abweichungsquadrate, wobei diejenige unabhängige Variable, die in einem gegebenen Schritt die größte Reduktion bewirkt, in die Regression aufgenommen wird.[4] Eine Folge von Variablen kann in die Regressionsfunktion "hineingezwungen" werden, andere Variable können aus der Gleichung gestrichen werden.

Die Kapazität von "STEPR" reicht bis zu 35 Variablen, d.h. bis zu maximal 34 unabhängigen Variablen, und bis zu 99.999

---

1) IBM Application Program SSP, a.a.O., S. 411.

2) Ebenda. S. 409.

3) Ebenda. S. 413.

4) Ebenda. (Sinngemäße Übersetzung des Verfassers).

Beobachtungen.[1]

Als Konstante kann in die Maschine eine Größe eingegeben werden, die den Anteil an der Summe der Abweichungsquadrate angibt, der als Grenze für die Aufnahme einer unabhängigen Variablen in die Regressionsgleichung festgesetzt wird. Wenn der Anteil in der Reduktion der Abweichungsquadrate, der auf eine einzubeziehende Variable entfällt, nicht größer als dieser festgesetzte Wert ist, wird diese Variable nicht in die Regression aufgenommen.[2] Die Größe dieses Wertes hängt natürlich vom Zweck der Analyse ab, ein Vorschlag wäre z.B. 0,01, d.h. 1 % der Gesamtsumme der Abweichungsquadrate. Möglich ist auch die Aufnahme von a l l e n unabhängigen Veränderlichen in die Regressionsgleichung, in diesem Fall wäre der einzugebende Wert gleich 0.

Die Rechenzeit für die Durchführung der Analyse auf der Anlage IBM 360/30 mit den schon genannten Randeinheiten liegt bei 45 Sekunden.[3]

Die Ausgabe der Maschine zeigt folgende Größen:[4]

1. Für alle Variablen:

   1.1 Mittelwerte,
   1.2 Standardabweichungen und
   1.3 Matrix der Korrelationskoeffizienten.

2. Für jeden Schritt der multiplen Regression:

   2.1 Summe der reduzierten Abweichungsquadrate,

---

1) Voraussetzung ist, daß die bereits in Fußnote 2 auf S. 263 genannten Bedingungen erfüllt sind. Andernfalls sind Modifikationen des Programms erforderlich.

2) IBM Application Program SSP, a.a.O., S. 413. Sinngemäße Übersetzung des Verfassers.

3) Ebenda. S. 417.

4) Ebenda. S. 415.

2.2 Reduzierter Anteil an der Gesamtabweichung,

2.3 Kumulierte Summe der reduzierten Abweichungsquadrate,

2.4 Kumulierter Anteil an der Gesamtabweichung,

2.5 Multipler Korrelationskoeffizient (verzerrt und unverzerrt),

2.6 F-Wert für die Durchführung einer Varianzanalyse,

2.7 Standardfehler der Schätzung (verzerrt und unverzerrt),

2.8 Regressionskoeffizienten einschließlich Regressionskonstante,

2.9 Standardfehler der Regressionskoeffizienten und

2.10 t-Werte zum Testen der Regressionskoeffizienten.

3. Tabelle der Residuen (wahlweise).

Mit Hilfe des Programms "STEPR" ist es möglich, für einen gegebenen Datensatz wirklich die "beste" - im Sinne der Methode der kleinsten Quadrate - Regressionsgleichung zu finden, da alle "wichtigen" Variablen in der Gleichung enthalten und alle "unwichtigen" Variablen außer Ansatz geblieben sind. Daneben bietet das Verfahren durch die sukzessive Hereinnahme von Variablen die Möglichkeit, mehrere Regressionsgleichungen mit unterschiedlicher Zahl von Variablen für betriebliche Zwecke zu nutzen.[1)]

Der an der Durchführung einer Regressions- und Korrelationsanalyse auf einer IBM-Anlage interessierte Leser findet weitere Angaben über IBM-Programme in den Veröffentlichungen der IBM-Rechenzentren unter dem Stichwort "Mathematische Statistik".[2)]

---

1) Vgl. hierzu auch die Darstellung der Vorteile der schrittweisen Regressionsanalyse auf S. 105 ff.

2) Vgl.: Korrelationsanalyse IR EIKO. IBM - Rechenzentren - Mathematische Statistik. Hrsg. von IBM Deutschland. Mai 1968. IBM Form 72 137 - 0.
Multiple Regressionsanalyse IR MURA. IBM - Rechenzentren - Mathematische Statistik. Hrsg. von IBM Deutschland. Mai 1968. IBM Form 72 140 - 0.

9.3.2 Siemens

In ähnlicher Weise wie die IBM hat die Firma Siemens ein Programmpaket der mathematischen Statistik unter der Bezeichnung "SIESTA" (SIEmens STAtistik) veröffentlicht, deren Programme für die Anlagen des Typs Siemens 4004/35, 45 und 55 vorgesehen sind und in FORTRAN IV geschrieben sind.[1)]

Im Rahmen dieses Paketes existieren vier Regressions- und Korrelationsprogramme, eine multiple lineare Regression (MLREG), eine lineare Regression mit Transformationen (REGRES), eine schrittweise lineare Regression mit Transformation (SLREG) und eine nichtlineare Regression mit eingebbarer Funktion (NLREG).[2)] Alle Programme bauen auf der Methode der kleinsten Quadrate auf.

Wie die Steuerprogramme der IBM sind auch die genannten Programme von Siemens aus verschiedenen Unterprogrammen zusammengesetzt, die in ähnlicher Weise arbeiten, so daß auf eine gesonderte Darstellung verzichtet werden kann.

Im folgenden soll eine kurze Charakterisierung der einzelnen Hauptprogramme genügen.[3)]

1. MLREG Kapazität: Insgesamt 40 Variable und 99.999 Beobachtungen.
   Kernspeicherbedarf: 43.500 Bytes.

---

1) Siemens-Statistik-System SIESTA für das Siemens-System 4004. Hrsg. von Siemens AG, o.O., Mai 1968. Kennzahl 543 12 1130 1. Statistik-System: Allgemeine Beschreibung für das Siemens-System 4004. Hrsg. von Siemens AG, o.O., vorl. Ausg. Oktober 1966. Best.-Nr. 2 - 2605 - 300.

2) Siemens-Statistik-System SIESTA, a.a.O., S. 0. Das Programm NLREG ist nicht als Programmband wie die übrigen Programme erhältlich, kann aber in FORTRAN IV auf Lochkarten bezogen werden.

3) Ebenda. S. 1/1 ff.

Ausgabe: Mittelwerte und Standardabweichungen der einzelnen Variablen; Korrelationskoeffizienten; Regressionskoeffizienten; Standardfehler der Regressionskoeffizienten; t-Werte; Regressionskonstante; multipler Korrelationskoeffizient; Standardfehler der Schätzung; Varianzanalyse für die multiple Regression; Tabelle der Residuen (wahlweise).

Die zur Durchführung der Analyse erforderliche Zeit wird von Siemens nicht angegeben, dürfte aber im Rahmen der von der IBM genannten Größenordnungen liegen.

Die Darstellung des Programms MLREG zeigt, daß es weitgehend mit dem Programm REGRE der IBM identisch ist und zu gleichen Ausgabedaten führt.

2. REGRES Die Besonderheit dieses Programms besteht in der Möglichkeit von Transformationen für die unabhängigen Variablen, z.B. Logarithmus zur Basis 10 oder e, e^x, $\sqrt{x}$ usw.[1)]

Kapazität: Insgesamt 9 Variable und 100 Beobachtungen.

Kernspeicherbedarf: 40.000 Bytes.

Ausgabe: Tabelle der Beobachtungen (Matrix aller Variablen); Mittelwerte und mittlere quadratische Abweichungen; einfache Korrela-

1) Vgl. Siemens-Statistik-System SIESTA, a.a.O., S. 2/2.

tionskoeffizienten und multipler Korrelationskoeffizient; Regressionskoeffizienten und -konstante; Standardfehler der Regressionskoeffizienten und -konstanten; t-Werte; F-Werte; y'_i-Werte und Residuen; Standardfehler der Schätzung.

3. SLREG Kapazität: 100 Variable, davon maximal 45 auszuwählende Variable, und 99.999 Beobachtungen.

Kernspeicherbedarf: 42.000 Bytes.

Auf eine besondere Darstellung dieses Programms kann verzichtet werden, da es weitgehend dem IBM-Programm STEPR entspricht und dem Analysierenden gleiche Möglichkeiten zur Berechnung der "besten" Regressionsgleichung bietet.[1)]

4. NLREG Dieses Regressions- und Korrelationsprogramm analysiert jedes nichtlineare Problem, das auf eine lineare Gleichung zurückgeführt werden kann.[2)]

Kapazität: Muß im Einzelfall anhand einer von Siemens angegebenen Formel errechnet werden.[3)]

Ausgabe: Identifikation und Anzahl der Parameter; maximal zugelassene Iterationen; zugelassener Fehler; Tabelle der Beobachtungswerte;

---

1) Daneben ist die Durchführung von "Faktorenanalysen" möglich. Hierauf kann im Rahmen dieser Arbeit jedoch nicht weiter eingegangen werden.

2) Siemens-Statistik-System SIESTA, a.a.O., S. 1/4.

3) Ebenda.

Iterationswerte: Summe der Fehlerquadrate und Näherungswerte für die Regressionskoeffizienten; Tabelle: Beobachteter und berechneter Wert der abhängigen Variablen und die Differenz (Residuen); multipler Korrelationskoeffizient; Standardfehler der Schätzung.

9.3.3 Bull General Electric

Auch für die Datenverarbeitungsanlagen der Firma Bull General Electric sind Programme zur Durchführung von Regressions- und Korrelationsanalysen entwickelt und als "Application Manual" veröffentlicht worden. Als Beispiel mag hier ein Programm für die Durchführung von multiplen linearen Regressionsanalysen dienen, mit dem auch die schrittweise Regressionsrechnung möglich ist und das für die Maschinen der Serie GE - 400 aufgestellt wurde.[1)]

Das multiple lineare Regressionsprogramm MLRP besteht aus drei aufeinander folgenden Programmphasen. In der ersten Phase (TRANSFORM) werden die Beobachtungswerte für die Variablen eingegeben und Transformationen für die unabhängigen Variablen geliefert. Die Phase 2 (GENERATE) berechnet die Mittelwerte und Standardabweichungen für alle Variablen und eine Matrix der einfachen Korrelationskoeffizienten. In der dritten Phase (ANALYSIS) erfolgt dann die Durchführung der eigentlichen Analyse mit der Berechnung der Regressionskoeffizienten, der Regressionskonstanten, der partiellen und des multiplen Korrelationskoeffizienten usw. Die schritt-

1) Multiple Linear Regression Program. GE - 400 Series. Application Manual. Hrsg. von Bull General Electric. O.O., Oktober 1965, durchges. Aufl. September 1969. CPB - 1158 C. Ref. 18.35.046 A. Auch dieses Programm basiert auf der Methode der kleinsten Quadrate.

weise Regressionsanalyse erfolgt ebenfalls in der dritten Phase ANALYSIS, wobei die Möglichkeit besteht, entweder a l l e unabhängigen Variablen - ohne Berücksichtigung ihrer t- oder F-Werte - in die Regressionsgleichung aufzunehmen (complete forcing) oder nur bestimmte ausgewählte Variable aufzunehmen (partial forcing), und zwar wiederum ohne Rücksicht auf ihre statistische Signifikanz.[1)]

Darüber hinaus besteht die Möglichkeit, diejenige unabhängige Variable in einem bestimmten Schritt in die Vorhersagegleichung aufzunehmen, die die größte Reduzierung der Summe der Abweichungsquadrate bewirkt, was der größten Zunahme des Bestimmtheitsmaßes entspricht.[2)] Oder man verlangt von dem Programm, jene unabhängigen Variablen hinzuzufügen, die einen bestimmten festgelegten F-Wert überschreiten. Jede Variable, deren berechneter F-Wert in einem Schritt diesen festgelegten Wert unterschreitet, wird als nichtsignifikant wieder aus der Vorhersagegleichung eliminiert. Der F-Wert wird dabei von dem Programm in jedem Schritt unter Berücksichtigung der Veränderungen in der Anzahl der Freiheitsgrade neu ermittelt. Dieses Verfahren wird auch als "forward addition and deletion" bezeichnet.[3)] Unter "backward deletion" versteht man eine Vorgehensweise, bei der die unabhängigen Variablen nach dem Kriterium der größten Reduktion in der Summe der Abweichungsquadrate in die Gleichung aufgenommen werden und bei der anschließend rückwärtsschreitend diejenigen Variablen eliminiert werden, die einen bestimmten festgelegten F-Wert nicht erreichen.

Das Programm bietet den Vorteil, daß, wenn im Fall der

---

1) Multiple Linear Regression Program. GE - 400 Series, a.a.O., S. 4 f.

2) Vgl. auch S. 265.

3) Multiple Linear Regression Program. GE - 400 Series, a.a.O., S. 5.

"forward addition and deletion" oder der "backward deletion" sich bei dem festgelegten Niveau keine Variable als signifikant erweist, automatisch eine Herabsetzung des Signifikanzniveaus (Wahl eines kleineren F-Wertes) erfolgt und dann die gleichen Rechenoperationen unter Verwendung dieses niedrigeren Wertes wiederholt werden.[1)]

Neben der Durchführung von linearen Regressionsanalysen kann das Programm MLRP auch für nichtlineare Gleichungen Verwendung finden, wenn sich die nichtlineare Form durch Transformationen auf eine lineare Form zurückführen läßt.[2)]

Eine weitere Besonderheit dieses Programms besteht in der Möglichkeit, die Regressionslinie durch den Ursprung des Koordinatensystems zu legen, d.h. eine bedingte Regressionsgleichung der Form

$$y' = b_1x_1 + b_2x_2 + \ldots + b_mx_m$$

zu verwenden.[3)]

Die Ausgabe des Programms in der dritten Phase besteht aus folgenden Größen, wobei alle Werte in jedem Fall für die endgültige Regressionsgleichung, auf Wunsch auch für jeden Regressionsschritt ausgedruckt werden:[4)]

Anzahl der Freiheitsgrade; Summe der Abweichungsquadrate; Standardfehler der Regressionsgeraden; multiples Be-

---

1) Multiple Linear Regression Program. GE - 400 Series, a.a.O., S. 5.

2) Ebenda. S. 7.

3) Ebenda. S. 8.

4) Zur Ausgabe in der zweiten Phase GENERATE vgl. S. 271. Auf die Bedeutung der "Zwischengleichungen", die während der Durchführung der schrittweisen Regressionsanalyse entstehen, wurde bereits auf S. 267 hingewiesen.

stimmtheitsmaß; partielle Regressionskoeffizienten; Standardfehler der Regressionskoeffizienten; F-Werte; partielle Korrelationskoeffizienten; Regressionskonstante; Schätzwerte y'_i; Residuen (wahlweise).

9.3.4 International Computers Limited (ICL)

Auch von der Firma International Computers Limited sind mehrere Programme zur Durchführung von Regressions- und Korrelationsanalysen entwickelt worden. Allerdings wurde auf Anfrage von ICL Deutschland in Düsseldorf mitgeteilt, daß diese statistischen Programme auf dem deutschen Markt noch nicht in nennenswertem Umfang eingesetzt werden. Es existieren jedoch Manuals über statistische Software, aus denen sich die erforderlichen Informationen zur Durchführung einer Regressionsanalyse entnehmen lassen.

Als Beispiel sei hier das Manual "Statistical Part 1" genannt,[1)] in dem Angaben zur Durchführung einer schrittweisen Regressionsanalyse auf Maschinen der Serie 1900 enthalten sind. Da sich dieses Programm - mit dem Namen "Multiple Regression Analysis" - weitgehend mit den bereits dargestellten Programmen der Firmen IBM, Siemens und Bull General Electric vergleichen läßt und auch eine ähnliche Datenausgabe liefert, kann hier auf eine gesonderte Darstellung verzichtet werden.

1) Scientific and technical software: Statistical, Part 1: Statistical analysis, system mark 1. ICL 1900 Series. Hrsg. vom Technical Publications Service der ICL (International Computers Limited). Technical Publication 3345/1. First edition. January 1966, reprinted March 1968.

9.3.5 Zuse

Für Datenverarbeitungsanlagen der Firma Zuse existieren ebenfalls Regressionsprogramme, darunter ein Programm für eine schrittweise Regressionsanalyse, das in ALGOL geschrieben wurde.[1] Auch hier soll die Erwähnung der Existenz dieser Programme genügen, da ihre Charakteristika bereits ausführlich dargestellt wurden.

9.3.6 Deutsches Rechenzentrum

Neben den genannten Programmen, die von Herstellerfirmen für jeweils spezielle Typen von Datenverarbeitungsanlagen entwickelt wurden, hält das Deutsche Rechenzentrum in Darmstadt eine umfangreiche Sammlung von Regressions- und Korrelationsprogrammen bereit, die aus verschiedenen Quellen gesammelt wurden.[2] Es sind

1. eigene Programme des Deutschen Rechenzentrums (DRZ) und stammen von Mitarbeitern oder Benutzern des DRZ oder wurden von ihnen für die Rechenanlage IBM 7094 des DRZ entwickelt.
2. Programme aus der Sammlung "BMD Biomedical Computer Programs", die die Health Sciences Computing Facility an der Universität von Kalifornien, Los Angeles, veröffentlicht hat.
3. Programme von der IBM-Benutzer-Organisation SHARE.

---

1) Multiple schrittweise Regressionsanalyse. Hrsg. von Zuse KG. Bad Hersfeld. O. J., Reg.-Nr. 2126.; Programmsystem für automatische Auswahlmethode bei Regressionsanalyse, (modifizierbarer GAUSS-JORDAN-Algorithmus) der Zuse Z 23. Hrsg. von Zuse KG., Bad Hersfeld, November 1964. Reg.-Nr. 1457.

2) Vgl. Programm-Information PI-4-2 des Deutschen Rechenzentrums (DRZ): Statistik-Programme. Darmstadt Januar 1968, verfaßt von F. GEBHARDT. S. 2 f.

Für die Programme des DRZ existieren ausführliche Programmbeschreibungen, die gegen eine Schutzgebühr beim Deutschen Rechenzentrum in Darmstadt bezogen werden können.[1] Die Programme selbst, die zum überwiegenden Teil in FORTRAN IV geschrieben sind, werden von den Benutzern des DRZ nicht benötigt, da sie wie beim "Scientific Subroutine Package" für das IBM-System 360 in binärer Form von einem speziellen Magnetband abgerufen werden können.[2] Hochschulrechenzentren können Kopien dieser Magnetbänder erhalten.[3] Auch in Lochkartenform können einzelne Programme bezogen werden, allerdings sind in diesem Fall zugehörige Unterprogramme gesondert bei der Programmbibliothek des DRZ zu bestellen. Über die Art der benötigten Unterprogramme geben die Programmbeschreibungen für die Hauptprogramme Auskunft.[4]

Die Programme, deren Standardversion für die Rechenanlage IBM 7094/I eingerichtet ist, können auf andere Rechenanlagen umgestellt werden. Für die Benutzung der weitverbreiteten Anlagen vom Typ IBM/360 oder für ähnlich konzipierte Anlagen werden vom DRZ besondere Hinweise gegeben.[5] Für die Benutzung der neu installierten Rechenanlage TR 440 der Firma AEG-Telefunken liegen zur Zeit noch keine Angaben zur Durchführung von Regressions- und Korrelationsanalysen vor.[6]

---

1) Programm-Information PI-32 des Deutschen Rechenzentrums (DRZ): Statistische Programme des DRZ, Teil A: Allgemeines. Darmstadt August 1969, verfaßt von F. GEBHARDT.; Programm-Information PI-33 des DRZ: Statistische Programme des DRZ, Teil B: Einzelbeschreibungen. Darmstadt August 1969, verfaßt von F. GEBHARDT.

2) PI-32 des DRZ, a.a.O., S. 7.

3) Ebenda.

4) Ebenda.

5) Ebenda. S. 14 f. u. S. 21 ff.

6) Ebenda. S. 24.

Aus der Vielzahl der DRZ-Programme, die für die Durchführung einer Regressions- und Korrelationsanalyse geschaffen wurden, sollen - um einen Überblick über ihre Charakteristika zu geben - die Programme "KOR 1" (Korrelationsprogramm 1) und "REGT" (Regressionsanalyse mit Tests) kurz dargestellt werden.

Das Korrelationsprogramm 1 (KOR 1)[1] ist ein in FORTRAN IV geschriebenes Hauptprogramm,[2] das der Berechnung einer Korrelations- oder Kovarianzmatrix dient. Die Zahl der Variablen darf höchstens 129 betragen. Es ist ein Kernspeicher mit 32 K erforderlich, allerdings dürften nach den Erfahrungen des DRZ bei etwa 60 Variablen auch 16 K ausreichen. Die Rechenzeit ermittelt sich nach der Formel

$$R = 0{,}02\ n\ k^2$$

bei großer Variablenzahl und ist in "msec" ausgedrückt.[3] Bei kleiner Variablenzahl überwiegt die Zeit zum Einlesen der Daten.

Die Ausgabe des Programms KOR 1 besteht in der Angabe der Mittelwerte und Varianzen für alle Variablen sowie in einer Matrix der Korrelationskoeffizienten bzw. Kovarianzen.

Das Regressionsprogramm REGT ist ebenfalls ein FORTRAN IV Hauptprogramm[4] mit verschiedenen Unterprogrammen zur Durchführung einer linearen schrittweisen Regression nach der

---

1) Zur ausführlichen Darstellung dieses Programms vgl. PI-33 des DRZ, a.a.O., S. 32 ff.

2) Die Autoren sind P. SCHNELL und F. GEBHARDT vom DRZ.

3) Die Symbole bedeuten: R = Rechenzeit in msec.; n = Zahl der Beobachtungen; k = Zahl der Variablen. Die Zeitangabe gilt für die Anlage IBM 7094/I.

4) Der Autor ist F. GEBHARDT.

Methode der kleinsten Quadrate, wobei wahlweise ein Test auf Normalverteilung der Residuen durchgeführt werden kann. Es können einzelne Regressionskoeffizienten vorgegeben werden, "insbesondere können die Regressionsflächen durch den Nullpunkt gelegt werden".[1)]

Die Zahl der Eingabevariablen darf höchstens 99 betragen, davon können 30 Variable in die Regressionsgleichung aufgenommen werden. Das Programm ist für einen 32 K Kernspeicher konzipiert, allerdings dürften bei einer Reduzierung auf etwa 10 bis 15 Eingabevariable, diese Zahl ist in aller Regel für kostenrechnerische Probleme ausreichend, 16 K genügen.[2)]

Der Zeitbedarf beträgt etwa 0,7 m in Sekunden für eine Regression mit m unabhängigen Variablen, außerdem für einen eventuell durchgeführten Normalitätstest weitere 0,015 n Sekunden, wobei n die Anzahl der Probanden (Beobachtungen) wiedergibt. Somit ergäbe sich für die Durchführung einer Analyse mit 30 unabhängigen Variablen - ohne Normalitätstest - ein Zeitbedarf von etwa 21 Sekunden.[3)]

Die Ausgabe des Programms REGT besteht aus folgenden Größen:[4)]

1. Für jede Variable, die in der Regression ist: Mittelwert; Standardabweichung; Regressionskoeffizient; Standardabweichung des Regressionskoeffizienten; t-Wert, um zu testen, ob der Regressionskoeffizient sig-

---

1) PI-33 des DRZ, a.a.O., S. 77.

2) Ebenda.

3) Die Angaben gelten für die Anlage IBM 7094/I.

4) PI-33 des DRZ, a.a.O., S. 79 ff.

nifikant von 0 verschieden ist, mit Angabe der Freiheitsgrade; partieller Korrelationskoeffizient einschließlich t-Wert; Irrtumswahrscheinlichkeit des t-Wertes; multipler Korrelationskoeffizient dieser Variablen mit allen anderen unabhängigen Variablen; Korrelation mit der abhängigen Variablen.

2. Für jede Regression:
   Zahl der Beobachtungen; Mittelwert der abhängigen Variablen; Standardabweichung des Mittelwerts; multiple Korrelation der abhängigen Variablen mit allen unabhängigen Variablen in der Regression; multiples Bestimmtheitsmaß mit F-Wert und Freiheitsgraden.

3. Werte für eingeschränkte und uneingeschränkte Hypothesen.[1)]

4. Angaben, wenn Test auf Normalität durchgeführt wird.

Die schrittweise Regression erfolgt bei dem Programm REGT in a b s t e i g e n d e r Reihenfolge, d.h. man beginnt "mit einem Satz unabhängiger Variabler, aus dem eine nach der anderen eliminiert wird, und zwar jeweils die mit dem absolut kleinsten t-Wert.[2)] Bei jedem Schritt wird eine komplette Regressionsanalyse durchgeführt. Es bleibt dann dem Fachwissenschaftler überlassen, welche Regressionsanalyse er aus dieser Serie verwenden will".[3)]

Neben diesen auch von anderen Programmen durchgeführten Be-

---

1) Vgl. hierzu S. 280 und PI-33 des DRZ, a.a.O., S. 80 ff.

2) Diese Berechnung entspricht der "backward deletion" des "Multiple Linear Regression Program" von Bull General Electric. Vgl. S. 272.

3) PI-33 des DRZ, a.a.O., S. 83.

rechnungen bietet das Programm REGT eine Besonderheit, die im Testen von "eingeschränkten Hypothesen" besteht. Es geht dabei um folgendes Problem:[1)]

"Es kommt häufig vor, daß man bei gleicher abhängiger Variabler mehrere Regressionen rechnen möchte, die sich in der Auswahl der unabhängigen Variablen unterscheiden. Enthält dabei die erste Regression die gleichen unabhängigen Variablen wie die zweite und noch einige zusätzliche, so kann man zugleich mit den üblichen Ergebnissen noch testen, ob diese zusätzlichen Variablen statistisch signifikant sind oder aber weggelassen werden dürfen. Dazu muß die erste Regression als 'uneingeschränkte Hypothese', die zweite als 'eingeschränkte Hypothese' formuliert werden.

Allgemein geht die eingeschränkte Hypothese durch beliebige lineare Restriktionen der Regressionskoeffizienten aus der uneingeschränkten Hypothese hervor, z.B. durch $a_1 = 0$, $a_2 = 47{,}11$, $a_3 = a_4$ $a_1 = 0$ bedeutet, daß Variable 1 in der eingeschränkten Hypothese nicht mehr vorkommt; $a_2 = 47{,}11$ bedeutet, daß Variable 2 in der eingeschränkten Hypothese den vorgegebenen Regressionskoeffizienten 47,11 erhält (...). Für die Einschränkung $a_3 = a_4$ schließlich benötigt man eine neue Variable (z.B. x_5), die gleich der Summe aus x_3 und x_4 ist (...); anstelle der beiden Variablen 3 und 4 enthält die eingeschränkte Hypothese dann nur die Variable 5."

Zum Testen dieser Einschränkungen dient ein F-Wert, der, wenn er signifikant ist, darauf hinweist, daß die vorgenommenen Einschränkungen nicht gerechtfertigt sind.[2)]

Neben diesen eigenen Programmen hält das Deutsche Rechenzentrum - wie bereits erwähnt - auch andere Programme zur Durchführung von Regressions- und Korrelationsanalysen bereit.

---

1) PI-33 des DRZ, a.a.O., S. 84.

2) Ebenda.

Die bedeutendste Gruppe dieser Programme, die meistens unverändert auf der Rechenanlage des DRZ benutzt werden können, bilden die BMD-Programme.[1] Sie sind in Fortran II und FAP geschrieben, darüber hinaus existiert eine FORTRAN IV - Version für Anlagen vom Typ IBM/360.[2]

Neben einigen einfachen Korrelationsprogrammen existieren zahlreiche Programme zur Durchführung von multiplen Regressionsanalysen. Besonders interessant sind dabei die folgenden Programme:[3]

BMD 01 R Simple Linear Regression

BMD 02 R Stepwise Regression
Das Programm berechnet nach dem üblichen Verfahren der größten Reduktion in der Fehlerquadratsumme eine Folge von Regressionsgleichungen, wobei in jedem Schritt eine Variable in die Gleichung aufgenommen wird. Variable mit nichtsignifikantem F-Testwert können unterdrückt werden.[4] Die Ausgabe entspricht weitgehend der Ausgabe der bisher genannten Programme für schrittweise Regressionsanalysen.

---

1) Vgl. S. 275.

2) PI-4-2 des DRZ, a.a.O., S. 3.

3) Vgl. PI-4-2 des DRZ, a.a.O., S. 18 ff. Hier muß eine kurze Aufzählung der BMD-Programme genügen. Wegen weitergehender Informationen wende man sich an das DRZ oder bestelle das Handbuch: W.J. DIXON (Hrsg.), BMD Biomedical Computer Programs. Health Sciences Computing Facility, University of California, Los Angeles 1965. Dieses Handbuch kann gegen Einsendung von U.S. Dollar 6,05 bezogen werden bei: UCLA Student Store, 308 Westwood Boulevard, Los Angeles, California 90024, U.S.A.

4) PI-4-2 des DRZ, a.a.O., S. 18.

BMD 05 R Polynomial Regression.
Die Regressionsfunktion ist ein Polynom k-ten Grades.

Darüber hinaus enthält die Statistik-Programmsammlung des DRZ mehrere Regressions- und Korrelationsprogramme der IBM-Benutzer-Organisation SHARE, die in FORTRAN II oder FORTRAN IV geschrieben sind und von verschiedenen Autoren stammen,[1)] teilweise auch von Mitarbeitern des DRZ für die Rechenanlage des DRZ modifiziert worden sind. Da diese Programme im wesentlichen die gleichen Ergebnisse liefern wie die bereits ausführlich dargestellten Programme, kann hier der Hinweis auf ihre Existenz und Bezugmöglichkeit genügen.

1) PI-4-2 des DRZ, a.a.O., S. 3.

10. Anhang II

EINIGE WICHTIGE TABELLEN UND NOMOGRAMME

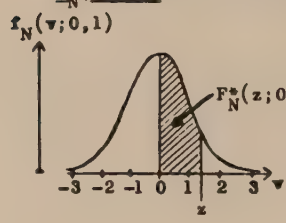
Die im folgenden wiedergegebenen statistischen Tabellen wurden entnommen aus der Sammlung "Statistische Tabellen" von J. BLEYMÜLLER.[1)]

---

1) J. BLEYMÜLLER: Statistische Tabellen. Zusammenstellung der wichtigsten theoretischen Verteilungen. Münster, Wintersemester 1968/69.

Die Standardnormalverteilung

Die aus der Verteilungsfunktion $F_N(z;0,1)$ abgeleitete Funktion $F_N^*(z;0,1)$



$F_N^*(z;0,1) = F_N(z;0,1) - 0,5 = W(0 < Z < z)$

$F_N^*(z;0,1)$ gibt also die Wahrscheinlichkeit an, die dem Intervall zwischen 0 und z zukommt.

Aus der Symmetrie der Normalverteilung folgt $F_N^*(z;0,1) = F_N^*(-z;0,1)$.

Ferner gilt: $F_N^*(0;0,1) = 0$.

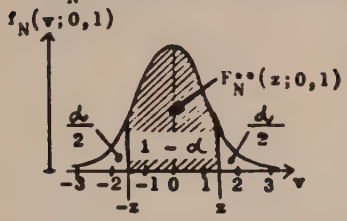
Ist die Variable X normalverteilt mit den Parametern μ und σ^2, dann ist die Variable $Z = \frac{X-\mu}{\sigma}$ normalverteilt mit den Parametern $\mu = 0$ und $\sigma^2 = 1$.

z	0	1	2	3	4	5	6	7	8	9
0,0	,0000	,0040	,0080	,0120	,0160	,0199	,0239	,0279	,0319	,0359
0,1	,0398	,0438	,0478	,0517	,0557	,0596	,0636	,0675	,0714	,0753
0,2	,0793	,0832	,0871	,0910	,0948	,0987	,1026	,1064	,1103	,1141
0,3	,1179	,1217	,1255	,1293	,1331	,1368	,1406	,1443	,1480	,1517
0,4	,1554	,1591	,1628	,1664	,1700	,1736	,1772	,1808	,1844	,1879
0,5	,1915	,1950	,1985	,2019	,2054	,2088	,2123	,2157	,2190	,2224
0,6	,2257	,2291	,2324	,2357	,2389	,2422	,2454	,2486	,2517	,2549
0,7	,2580	,2611	,2642	,2673	,2704	,2734	,2764	,2794	,2823	,2852
0,8	,2881	,2910	,2939	,2967	,2995	,3023	,3051	,3078	,3106	,3133
0,9	,3159	,3186	,3212	,3238	,3264	,3289	,3315	,3340	,3365	,3389
1,0	,3413	,3438	,3461	,3485	,3508	,3531	,3554	,3577	,3599	,3621
1,1	,3643	,3665	,3686	,3708	,3729	,3749	,3770	,3790	,3810	,3830
1,2	,3849	,3869	,3888	,3907	,3925	,3944	,3962	,3980	,3997	,4015
1,3	,4032	,4049	,4066	,4082	,4099	,4115	,4131	,4147	,4162	,4177
1,4	,4192	,4207	,4222	,4236	,4251	,4265	,4279	,4292	,4306	,4319
1,5	,4332	,4345	,4357	,4370	,4382	,4394	,4406	,4418	,4429	,4441
1,6	,4452	,4463	,4474	,4484	,4495	,4505	,4515	,4525	,4535	,4545
1,7	,4554	,4564	,4573	,4582	,4591	,4599	,4608	,4616	,4625	,4633
1,8	,4641	,4649	,4656	,4664	,4671	,4678	,4686	,4693	,4699	,4706
1,9	,4713	,4719	,4726	,4732	,4738	,4744	,4750	,4756	,4761	,4767
2,0	,4772	,4778	,4783	,4788	,4793	,4798	,4803	,4808	,4812	,4817
2,1	,4821	,4826	,4830	,4834	,4838	,4842	,4846	,4850	,4854	,4857
2,2	,4861	,4864	,4868	,4871	,4875	,4878	,4881	,4884	,4887	,4890
2,3	,4893	,4896	,4898	,4901	,4904	,4906	,4909	,4911	,4913	,4916
2,4	,4918	,4920	,4922	,4925	,4927	,4929	,4931	,4932	,4934	,4936
2,5	,4938	,4940	,4941	,4943	,4945	,4946	,4948	,4949	,4951	,4952
2,6	,4953	,4955	,4956	,4957	,4959	,4960	,4961	,4962	,4963	,4964
2,7	,4965	,4966	,4967	,4968	,4969	,4970	,4971	,4972	,4973	,4974
2,8	,4974	,4975	,4976	,4977	,4977	,4978	,4979	,4979	,4980	,4981
2,9	,4981	,4982	,4982	,4983	,4984	,4984	,4985	,4985	,4986	,4986
3,0	,4987	,4987	,4987	,4988	,4988	,4989	,4989	,4989	,4990	,4990

Tabelle 1*: Standardnormalverteilung (einseitig)

Die Standardnormalverteilung

Die aus der Verteilungsfunktion $F_N(z;0,1)$ abgeleitete Funktion $F_N^{**}(z;0,1)$



$F_N^{**}(z;0,1) = F_N(z;0,1) - F_N(-z;0,1) = W(-z < Z < z)$

Da $F_N(-z) = 1 - F_N(z)$, gilt

$F_N^{**}(z;0,1) = 2F(z;0,1) - 1.$

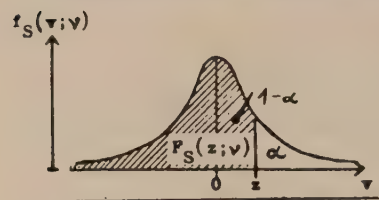
Ist die Variable X normalverteilt mit den Parametern μ und σ^2, dann ist die Variable $Z = \frac{X-\mu}{\sigma}$ normalverteilt mit den Parametern $\mu = 0$ und $\sigma^2 = 1$.

z	0	1	2	3	4	5	6	7	8	9
0,0	,0000	,0080	,0160	,0239	,0319	,0399	,0478	,0558	,0638	,0717
0,1	,0797	,0876	,0955	,1034	,1113	,1192	,1271	,1350	,1428	,1507
0,2	,1585	,1663	,1741	,1819	,1897	,1974	,2051	,2128	,2205	,2282
0,3	,2358	,2434	,2510	,2586	,2661	,2737	,2812	,2886	,2961	,3035
0,4	,3108	,3182	,3255	,3328	,3401	,3473	,3545	,3616	,3688	,3759
0,5	,3829	,3899	,3969	,4039	,4108	,4177	,4245	,4313	,4381	,4448
0,6	,4515	,4581	,4647	,4713	,4778	,4843	,4907	,4971	,5035	,5098
0,7	,5161	,5223	,5285	,5346	,5407	,5467	,5527	,5587	,5646	,5705
0,8	,5763	,5821	,5878	,5935	,5991	,6047	,6102	,6157	,6211	,6265
0,9	,6319	,6372	,6424	,6476	,6528	,6579	,6629	,6680	,6729	,6778
1,0	,6827	,6875	,6923	,6970	,7017	,7063	,7109	,7154	,7199	,7243
1,1	,7287	,7330	,7373	,7415	,7457	,7499	,7540	,7580	,7620	,7660
1,2	,7699	,7737	,7775	,7813	,7850	,7887	,7923	,7959	,7995	,8029
1,3	,8064	,8098	,8132	,8165	,8198	,8230	,8262	,8293	,8324	,8355
1,4	,8385	,8415	,8444	,8473	,8501	,8529	,8557	,8584	,8611	,8638
1,5	,8664	,8690	,8715	,8740	,8764	,8789	,8812	,8836	,8859	,8882
1,6	,8904	,8926	,8948	,8969	,8990	,9011	,9031	,9051	,9070	,9090
1,7	,9109	,9127	,9146	,9164	,9181	,9199	,9216	,9233	,9249	,9265
1,8	,9281	,9297	,9312	,9328	,9342	,9357	,9371	,9385	,9399	,9412
1,9	,9426	,9439	,9451	,9464	,9476	,9488	,9500	,9512	,9523	,9534
2,0	,9545	,9556	,9566	,9576	,9586	,9596	,9606	,9615	,9625	,9634
2,1	,9643	,9651	,9660	,9668	,9676	,9684	,9692	,9700	,9707	,9715
2,2	,9722	,9729	,9736	,9743	,9749	,9756	,9762	,9768	,9774	,9780
2,3	,9786	,9791	,9797	,9802	,9807	,9812	,9817	,9822	,9827	,9832
2,4	,9836	,9840	,9845	,9849	,9853	,9857	,9861	,9865	,9869	,9872
2,5	,9876	,9879	,9883	,9886	,9889	,9892	,9895	,9898	,9901	,9904
2,6	,9907	,9909	,9912	,9915	,9917	,9920	,9922	,9924	,9926	,9929
2,7	,9931	,9933	,9935	,9937	,9939	,9940	,9942	,9944	,9946	,9947
2,8	,9949	,9950	,9952	,9953	,9955	,9956	,9958	,9959	,9960	,9961
2,9	,9963	,9964	,9965	,9966	,9967	,9968	,9969	,9970	,9971	,9972
3,0	,9973	,9974	,9975	,9976	,9976	,9977	,9978	,9979	,9979	,9980

Tabelle 2*: Standardnormalverteilung (zweiseitig)

Die Studentverteilung

Die Werte von z, die gegebenen Werten der Verteilungsfunktion der Studentverteilung (t-Verteilung), also $P_S(z;\nu)$, entsprechen



$$P_S(z;\nu) = W(-\infty < Z \leq z) = 1 - \alpha$$

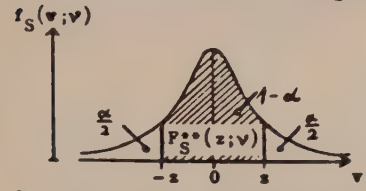
$$- P_S(z;\nu) = W(Z > z) = \alpha$$

ν	$P_S(z;\nu) = W(-\infty<Z \leq z) = 1-\alpha =$							
	0,75	0,90	0,95	0,975	0,99	0,995	0,999	0,9995
1	1,000	3,078	6,314	12,706	31,821	63,657	318,3	636,619
2	0,816	1,886	2,920	4,303	6,965	9,925	22,33	31,598
3	0,765	1,638	2,353	3,182	4,541	5,841	10,21	12,941
4	0,741	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	0,727	1,476	2,015	2,571	3,365	4,032	5,893	6,869
6	0,718	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	0,711	1,415	1,895	2,365	2,998	3,499	4,785	5,408
8	0,706	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	0,703	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	0,700	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	0,697	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	0,695	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	0,694	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	0,692	1,345	1,761	2,145	2,624	2,977	3,787	4,140
15	0,691	1,341	1,753	2,131	2,602	2,947	3,733	4,073
16	0,690	1,337	1,746	2,120	2,583	2,921	3,686	4,015
17	0,689	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	0,688	1,330	1,734	2,101	2,552	2,878	3,610	3,922
19	0,688	1,328	1,729	2,093	2,539	2,861	3,579	3,883
20	0,687	1,325	1,725	2,086	2,528	2,845	3,552	3,850
21	0,686	1,323	1,721	2,080	2,518	2,831	3,527	3,819
22	0,686	1,321	1,717	2,074	2,508	2,819	3,505	3,792
23	0,685	1,319	1,714	2,069	2,500	2,807	3,485	3,768
24	0,685	1,318	1,711	2,064	2,492	2,797	3,467	3,745
25	0,684	1,316	1,708	2,060	2,485	2,787	3,450	3,725
26	0,684	1,315	1,706	2,056	2,479	2,779	3,435	3,707
27	0,684	1,314	1,703	2,052	2,473	2,771	3,421	3,690
28	0,683	1,313	1,701	2,048	2,467	2,763	3,408	3,674
29	0,683	1,311	1,699	2,045	2,462	2,756	3,396	3,659
30	0,683	1,310	1,697	2,042	2,457	2,750	3,385	3,646
40	0,681	1,303	1,684	2,021	2,423	2,704	3,307	3,551
50	0,679	1,299	1,676	2,009	2,403	2,678	3,261	3,496
100	0,677	1,290	1,660	1,984	2,364	2,626	3,174	3,390
200	0,676	1,286	1,652	1,972	2,345	2,601	3,131	3,340
∞	0,675	1,282	1,645	1,960	2,326	2,576	3,090	3,291

Tabelle 3*: Studentverteilung (einseitig)

Die Studentverteilung

Die Werte von z, die gegebenen Werten der aus der Verteilungsfunktion der Studentverteilung (t-Verteilung), also aus $F_S(z;\nu)$, abgeleiteten Funktion $F_S^{**}(z;\nu)$ entsprechen



$$F_S^{**}(z;\nu) = W(-z < Z < z) = 1 - \alpha$$

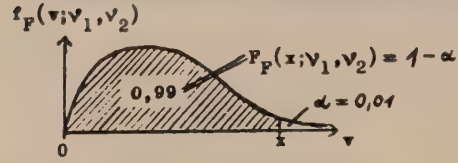
$$1 - F_S^{**}(z;\nu) = \alpha$$

ν	$F_S^{**}(z;\nu) = W(-z < Z < z) = 1-\alpha =$							
	0,50	0,80	0,90	0,95	0,98	0,99	0,998	0,999
1	1,000	3,078	6,314	12,706	31,821	63,657	318,3	636,619
2	0,816	1,886	2,920	4,303	6,965	9,925	22,33	31,598
3	0,765	1,638	2,353	3,182	4,541	5,841	10,21	12,941
4	0,741	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	0,727	1,476	2,015	2,571	3,365	4,032	5,893	6,869
6	0,718	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	0,711	1,415	1,895	2,365	2,998	3,499	4,785	5,408
8	0,706	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	0,703	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	0,700	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	0,697	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	0,695	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	0,894	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	0,692	1,345	1,761	2,145	2,624	2,977	3,787	4,140
15	0,691	1,341	1,753	2,131	2,602	2,947	3,733	4,073
16	0,690	1,337	1,746	2,120	2,583	2,921	3,686	4,015
17	0,689	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	0,688	1,330	1,734	2,101	2,552	2,878	3,610	3,922
19	0,688	1,328	1,729	2,093	2,539	2,861	3,579	3,883
20	0,687	1,325	1,725	2,086	2,528	2,845	3,552	3,850
21	0,686	1,323	1,721	2,080	2,518	2,831	3,527	3,819
22	0,686	1,321	1,717	2,074	2,508	2,819	3,505	3,792
23	0,685	1,319	1,714	2,069	2,500	2,807	3,485	3,768
24	0,685	1,318	1,711	2,064	2,492	2,797	3,467	3,745
25	0,684	1,316	1,708	2,060	2,485	2,787	3,450	3,725
26	0,684	1,315	1,706	2,056	2,479	2,779	3,435	3,707
27	0,684	1,314	1,703	2,052	2,473	2,771	3,421	3,690
28	0,683	1,313	1,701	2,048	2,467	2,763	3,408	3,674
29	0,683	1,311	1,699	2,045	2,462	2,756	3,396	3,659
30	0,683	1,310	1,697	2,042	2,457	2,750	3,385	3,646
40	0,681	1,303	1,684	2,021	2,423	2,704	3,307	3,551
50	0,679	1,299	1,676	2,009	2,403	2,678	3,261	3,496
100	0,677	1,290	1,660	1,984	2,364	2,626	3,174	3,390
200	0,676	1,286	1,652	1,972	2,345	2,601	3,131	3,340
∞	0,675	1,282	1,645	1,960	2,326	2,576	3,090	3,291

Tabelle 4*: Studentverteilung (zweiseitig)

Die F-Verteilung

Die Werte von x, für die die Verteilungsfunktion der F-Verteilung mit (ν_1,ν_2) Freiheitsgraden, also $F_F(x;\nu_1,\nu_2)$, den Wert 0,99 hat



$$F_F(x;\nu_1,\nu_2) = W(0 < X \leq x) = 0,99$$

ν_2 \ ν_1	1	2	3	4	5	6	8	12	24	∞
1	4052	4999	5403	5625	5764	5859	5981	6106	6235	6366
2	98,49	99,01	99,17	99,25	99,30	99,33	99,36	99,42	99,46	99,50
3	34,12	30,81	29,46	28,71	28,24	27,91	27,49	27,05	26,60	26,12
4	21,20	18,00	16,69	15,98	15,52	15,21	14,80	14,37	13,93	13,46
5	16,26	13,27	12,06	11,39	11,00	10,67	10,27	9,89	9,47	9,02
6	13,74	10,92	9,78	9,15	8,75	8,47	8,10	7,72	7,31	6,88
7	12,25	9,55	8,45	7,85	7,46	7,19	6,84	6,47	6,07	5,65
8	11,26	8,65	7,59	7,01	6,63	6,37	6,03	5,67	5,28	4,86
9	10,56	8,02	6,99	6,42	6,06	5,80	5,47	5,11	4,73	4,31
10	10,04	7,56	6,55	5,99	5,64	5,39	5,06	4,71	4,33	3,91
11	9,64	7,20	6,21	5,67	5,31	5,07	4,74	4,39	4,02	3,60
12	9,33	6,93	5,95	5,41	5,06	4,82	4,50	4,16	3,78	3,36
13	9,07	6,70	5,74	5,21	4,86	4,62	4,30	3,96	3,59	3,17
14	8,86	6,51	5,56	5,04	4,69	4,46	4,14	3,80	3,43	3,00
15	8,68	6,36	5,42	4,89	4,56	4,32	4,00	3,67	3,29	2,87
16	8,53	6,23	5,29	4,77	4,44	4,20	3,89	3,55	3,18	2,75
17	8,40	6,11	5,18	4,67	4,34	4,10	3,79	3,46	3,08	2,65
18	8,29	6,01	5,09	4,58	4,25	4,01	3,71	3,37	3,00	2,57
19	8,18	5,93	5,01	4,50	4,17	3,94	3,63	3,30	2,92	2,49
20	8,10	5,85	4,94	4,43	4,10	3,87	3,56	3,23	2,86	2,42
21	8,02	5,78	4,87	4,37	4,04	3,81	3,51	3,17	2,80	2,36
22	7,95	5,72	4,82	4,31	3,99	3,76	3,45	3,12	2,75	2,31
23	7,88	5,66	4,76	4,26	3,94	3,71	3,41	3,07	2,70	2,26
24	7,82	5,61	4,72	4,22	3,90	3,67	3,36	3,03	2,66	2,21
25	7,77	5,57	4,68	4,18	3,85	3,63	3,32	2,99	2,62	2,17
26	7,72	5,53	4,64	4,14	3,82	3,59	3,29	2,96	2,58	2,13
27	7,68	5,49	4,60	4,11	3,78	3,56	3,26	2,93	2,55	2,10
28	7,64	5,45	4,57	4,07	3,75	3,53	3,23	2,90	2,52	2,06
29	7,60	5,42	4,54	4,04	3,73	3,50	3,20	2,87	2,49	2,03
30	7,56	5,39	4,51	4,02	3,70	3,47	3,17	2,84	2,47	2,01
40	7,31	5,18	4,31	3,83	3,51	3,29	2,99	2,66	2,29	1,80
50	7,17	5,06	4,20	3,72	3,41	3,19	2,89	2,56	2,18	1,68
60	7,07	4,98	4,13	3,65	3,34	3,12	2,82	2,50	2,12	1,60
80	6,96	4,88	4,04	3,56	3,25	3,04	2,74	2,42	2,03	1,49
100	6,89	4,82	3,98	3,51	3,21	2,99	2,69	2,37	1,98	1,43
200	6,75	4,71	3,88	3,41	3,11	2,89	2,60	2,27	1,89	1,28
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,51	2,18	1,79	1,00

Tabelle 5*: F-Verteilung bei einem Konfidenzniveau von 99 %

Die Transformation von r in z

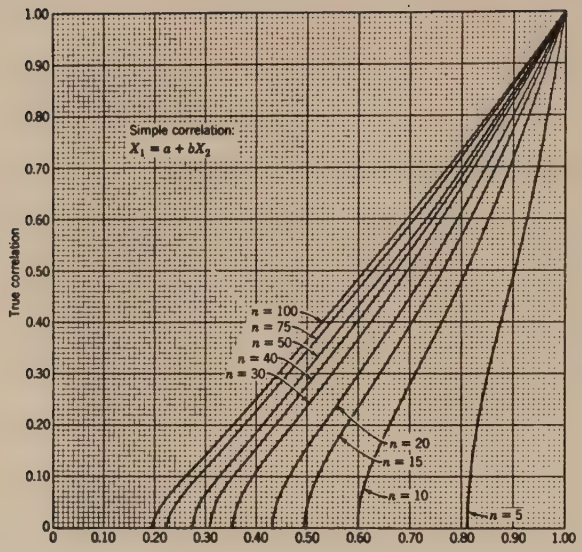
Die Stichprobenverteilung der Werte $z = \frac{1}{2} \ln \frac{1+r}{1-r} = 1{,}1513 \lg \frac{1+r}{1-r}$ nähert sich bei wachsendem Stichprobenumfang n schnell der Normalverteilung mit den Parametern $\mu_Z \approx \frac{1}{2} \ln \frac{1+\varrho}{1-\varrho} + \frac{\varrho}{2(n-1)}$ und $\sigma_Z^2 = \frac{1}{n-3}$.

Die z-Werte sind unten für $r \geq 0$ tabelliert;

für $r < 0$ gilt $-z = \frac{1}{2} \ln \frac{1+|r|}{1-|r|}$.

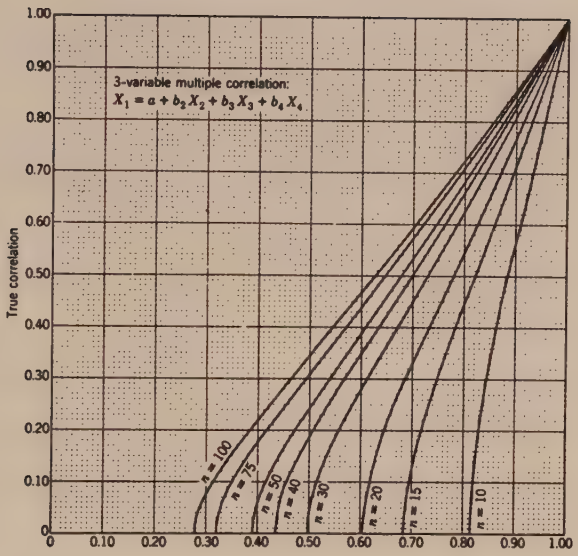
r	0	1	2	3	4	5	6	7	8	9
0,0	0,000	0,010	0,020	0,030	0,040	0,050	0,060	0,070	0,080	0,090
0,1	0,100	0,110	0,121	0,131	0,141	0,151	0,161	0,172	0,182	0,192
0,2	0,203	0,213	0,224	0,234	0,245	0,255	0,266	0,277	0,288	0,299
0,3	0,310	0,321	0,332	0,343	0,354	0,365	0,377	0,388	0,400	0,412
0,4	0,424	0,436	0,448	0,460	0,472	0,485	0,497	0,510	0,523	0,536
0,5	0,549	0,563	0,576	0,590	0,604	0,618	0,633	0,648	0,662	0,678
0,6	0,693	0,709	0,725	0,741	0,758	0,775	0,793	0,811	0,829	0,848
0,7	0,867	0,887	0,908	0,929	0,950	0,973	0,996	1,020	1,045	1,071
0,8	1,099	1,127	1,157	1,188	1,221	1,256	1,293	1,333	1,376	1,422
0,9	1,472	1,528	1,589	1,658	1,738	1,832	1,946	2,092	2,298	2,647

Tabelle 6*: z-Transformation nach FISHER



Nomogramm 7*: Nomogramm zur Ermittlung der Mindesthöhe des einfachen Korrelationskoeffizienten ϱ_{yx} bei unterschiedlichen Stichprobenkorrelationskoeffizienten r_{yx} und unterschiedlichem Stichprobenumfang n. Der Sicherheitsgrad der Aussage beträgt 95 %.[1)]

1) Entnommen aus M. EZEKIEL u. K.A. FOX, a.a.O., S. 294, Fig. 17.2. Die Wiedergabe dieses und des folgenden Nomogramms geschieht mit freundlicher Genehmigung des Verlags John Wiley & Sons, Inc., New York, U.S.A.



Nomogramm 8*: Nomogramm zur Ermittlung der Mindesthöhe des multiplen Korrelationskoeffizienten der Grundgesamtheit $\varrho_{y.123}$ bei unterschiedlichen Stichprobenkorrelationskoeffizienten $r_{y.123}$ und unterschiedlichem Stichprobenumfang n. Der Sicherheitsgrad der Aussage beträgt 95 %.[1)]

1) Entnommen aus M. EZEKIEL u. K.A. FOX, a.a.O., S. 296, Fig. 17.3.

VERZEICHNIS DER BENUTZTEN SYMBOLE

a_{yx} Regressionskonstante der Stichprobenregressionsfunktion bei einfacher Regression.

α_{yx} Regressionskonstante der Regressionsfunktion der Grundgesamtheit bei einfacher Regression.

$a_{y \cdot 12...m}$ Regressionskonstante der Stichprobe bei multipler Regression.

b_{yx} Regressionskoeffizient der Stichprobenregressionsfunktion bei einfacher Regression.

β_{yx} Regressionskoeffizient der Regressionsfunktion der Grundgesamtheit bei einfacher Regression.

$b_{y1 \cdot 23...m}$ partieller Regressionskoeffizient der Stichprobe bei multipler Regression.

$\beta_{y1 \cdot 23...m}$ partieller Regressionskoeffizient der Grundgesamtheit bei multipler Regression.

c_{yx} Regressionskoeffizient der Stichprobe (er findet Verwendung bei nichtlinearer Einfachregression).

d_{yx} einfaches Bestimmtheitsmaß in der Stichprobe.

$d_{y \cdot 12...m}$ multiples Bestimmtheitsmaß in der Stichprobe.

$\delta_{y \cdot 12...m}$ multiples Bestimmtheitsmaß der Grundgesamtheit.

e_i	$e_i = (y_i - y'_i)$, Abweichung zwischen beobachtetem Wert der abhängigen Variablen und Stichprobenregreßwert (für i = 1, 2, ..., n).
k	Zahl der unabhängigen Variablen in der Regressionsgleichung (für k = 1, 2, ..., m).
n	Stichprobenumfang.
ν	Zahl der Freiheitsgrade.
r_{yx}	einfacher Korrelationskoeffizient in der Stichprobe.
ϱ_{yx}	einfacher Korrelationskoeffizient in der Grundgesamtheit.
$r_{y\cdot 12\ldots m}$	multipler Korrelationskoeffizient in der Stichprobe.
$\varrho_{y\cdot 12\ldots m}$	multipler Korrelationskoeffizient in der Grundgesamtheit.
$r_{y1\cdot 23\ldots m}$	partieller Korrelationskoeffizient in der Stichprobe.
$\varrho_{y1\cdot 23\ldots m}$	partieller Korrelationskoeffizient in der Grundgesamtheit.
s_x	Standardabweichung der x-Werte in der Stichprobe.
σ_x	Standardabweichung der X-Werte in der Grundgesamtheit.
$\hat{\sigma}_x$	unverzerrter Schätzwert für die Standardabweichung der Grundgesamtheit.

$s_{y.x}$ Standardabweichung der Residuen (Standardfehler der Schätzung) in der Stichprobe.

$\sigma_{y.x}$ Standardfehler der Schätzung in der Grundgesamtheit.

$\hat{\sigma}_{y.x}$ unverzerrter Schätzwert für den Standardfehler der Schätzung in der Grundgesamtheit.

$s_{b_{yx}}$ Standardfehler des einfachen Regressionskoeffizienten in der Stichprobe.

$\hat{\sigma}_{b_{yx}}$ unverzerrter Schätzwert für den Standardfehler des einfachen Regressionskoeffizienten der Grundgesamtheit.

$s_{b_{y1.23...m}}$ Standardfehler des partiellen Regressionskoeffizienten in der Stichprobe.

$\hat{\sigma}_{b_{y1.23...m}}$ unverzerrter Schätzwert für den Standardfehler des partiellen Regressionskoeffizienten der Grundgesamtheit.

$s_{y'}$ Standardfehler der y-Koordinate eines Punktes auf der Regressionsgeraden (Standardfehler der Regressionsgeraden) in der Stichprobe.

$\hat{\sigma}_{y'}$ unverzerrter Schätzwert für den Standardfehler der Regressionsgeraden in der Grundgesamtheit.

s_F Standardfehler einer individuellen Prognose in der Stichprobe.

$\hat{\sigma}_F$ unverzerrter Schätzwert für den Standardfehler einer individuellen Prognose in der Grundgesamtheit.

x_i Stichprobenbeobachtungswert der unabhängigen Veränderlichen (für i = 1, 2, ..., n).

X_i entsprechender Wert der Grundgesamtheit.

x_{ik} Stichprobenbeobachtungswert der unabhängigen Veränderlichen k bei multiplen Regressionsanalysen (für k = 1, 2, ..., m).

$\bar{x}$ arithmetisches Mittel aus den Werten der unabhängigen Veränderlichen.

y_i Stichprobenbeobachtungswert der abhängigen Veränderlichen (für i = 1, 2, ..., n).

Y_i entsprechender Wert der Grundgesamtheit.

$\bar{y}$ arithmetisches Mittel aus den Werten der abhängigen Veränderlichen.

y'_i aus der Stichprobenregressionsgleichung geschätzter Wert für die abhängige Variable (Regreßwert).

Y'_i "wahrer" Regreßwert der Grundgesamtheit.

L I T E R A T U R V E R Z E I C H N I S

1. Bücher

ALLEN, Roy George Douglas: Statistik für Volkswirte. Aus dem Englischen übertragen von Wolfgang FÖRSTER. 2., unveränd. Aufl. Tübingen 1964.

ANDERSON, Oskar: Probleme der statistischen Methodenlehre in den Sozialwissenschaften. 5., unveränd. Aufl. Würzburg 1965.

BÖHM, Hans-Hermann und WILLE, Friedrich: Deckungsbeitragsrechnung und Programmoptimierung. 2., völlig überarb. u. erw. Aufl. des Werkes "Direct Costing und Programmplanung". München 1965.

BONNET, Karsten: Aufbau und Anwendungsmöglichkeiten des Programms "REGRE" (Regressions- und Korrelationsanalyse) aus der Programmbibliothek zum System IBM-360. Unveröffentl. Dipl.-Arbeit am Institut für Ökonometrie und Wirtschaftsstatistik der Universität Münster. Münster 1969.

BUSSMANN, Karl Ferdinand: Industrielles Rechnungswesen. Stuttgart 1963.

CROXTON, Frederick E. und COWDEN, Dudley J.: Applied General Statistics. Second edition. London (1965).

DRAPER, N.R. und SMITH, H.: Applied Regression Analysis. Third printing. New York, London und Sydney 1967.

EZEKIEL, Mordecai und FOX, Karl A.: Methods of Correlation and Regression Analysis. Linear and Curvilinear. Third edition, fifth printing. New York, London und Sydney 1966.

FISCHER, Johannes, HESS, Otto und SEEBAUER, Georg: Buchführung und Kostenrechnung. 5. Aufl. Leipzig 1942.

FÖRSTER, Erhard und EGERMAYER, Frantisek: Korrelations- und Regressionsanalyse. Ein Leitfaden für Ökonomen. Berlin (Ost) 1966.

FUNKE, Hermann, MELLEROWICZ, Konrad und ABROMEIT, Hans-Günther: Grundfragen und Technik der Betriebsabrechnung. 3. Aufl. Freiburg i. Br. 1962.

GOLDBERGER, Arthur S.: Topics in Regression Analysis. New York und London 1968.

GOULDEN, Cyril H.: Methods of Statistical Analysis. Second edition. New York und London (1952).

GRÖNER, Gerhard: Ein Beitrag zur Anwendung der Mehrfachkorrelation auf wirtschaftliche Reihen. Diss. an der Wirtschaftshochschule Mannheim. Mannheim 1957.

GUTENBERG, Erich: Einführung in die Betriebswirtschaftslehre. (Die Wirtschaftswissenschaften. Reihe A, Betriebswirtschaftslehre, Nr. 1, Hrsg. Erich GUTENBERG.) Wiesbaden 1958.

GUTENBERG, Erich: Grundlagen der Betriebswirtschaftslehre. Band 1: Die Produktion. 8./9. Aufl. Berlin, Göttingen und Heidelberg 1963.

GUTENBERG, Erich: Grundlagen der Betriebswirtschaftslehre. Band 1: Die Produktion. 17. Aufl. Berlin, Heidelberg und New York 1970.

HALD, Anders: Statistical Theory with Engineering Applications. Fifth printing. New York und London 1962.

HEINEN, Edmund: Betriebswirtschaftliche Kostenlehre. Kostentheorie und Kostenentscheidungen. 3., verb. Aufl. Wiesbaden 1970.

HELLWIG, Zdzislaw: Linear Regression and its Application to Economics. Translated from the Polish by J. STADLER. Oxford, London, New York, Paris und Warszawa 1963.

HENZEL, Friedrich: Kosten und Leistung. 3., unveränd. Aufl. der Kostenanalyse. Stuttgart 1957.

HENZEL, Friedrich: Die Kostenrechnung. 4., erw. u. veränd. Aufl. Essen 1964.

KÄFER, Karl: Standardkostenrechnung. Lehr- und Handbuch. 2., neubearb. u. stark erw. Aufl. Stuttgart 1964.

KELLERER, Hans: Statistik im modernen Wirtschafts- und Sozialleben. (Band 103/104 von Rowohlts Deutscher Enzyklopädie, Sachgebiet Wirtschaftswissenschaften, Hrsg. ERNESTO GRASSI). O.O. (1963).

KENDALL, Maurice George: The Advanced Theory of Statistics. Volume 1. Third edition. London 1947.

KILGER, Wolfgang: Flexible Plankostenrechnung. Theorie und Praxis der Grenzplankostenrechnung und Deckungsbeitragsrechnung. (Band 31 der Veröffentlichungen der SCHMALENBACH-Gesellschaft). 4. Aufl. Köln und Opladen 1970.

KLINGST, Anna: Einführung in die Matrizenrechnung. In: Anwendungen der Matrizenrechnung auf wirtschaftliche und statistische Probleme. 3. Aufl. Würzburg und Wien 1966.

KOCH, Helmut: Grundprobleme der Kostenrechnung. Köln und Opladen 1966.

KOSIOL, Erich: Kalkulatorische Buchhaltung (Betriebsbuchhaltung). Systematische Darstellung der Betriebsabrechnung und der kurzfristigen Erfolgsrechnung. 5., durchges. Aufl. Wiesbaden 1953.

KOSIOL, Erich (Hrsg.): Plankostenrechnung als Instrument moderner Unternehmungsführung. Erhebungen und Studien zur grundsätzlichen Problematik. (Veröffentlichungen des Betriebswirtschaftlichen Instituts für Industrieforschung der Freien Universität Berlin). Berlin 1956.

LASSMANN, Gert: Die Kosten- und Erlösrechnung als Instrument der Planung und Kontrolle in Industriebetrieben. Betriebswirtschaftliches Institut der Eisenhüttenindustrie. Düsseldorf 1968.

LEHMANN, Max Rudolf: Industriekalkulation. (Betriebswirtschaftliche Bibliothek. Hrsg. Wilhelm HASENACK. Reihe A/VII.) 5. Aufl. Essen 1964.

LINDER, Arthur: Statistische Methoden. Für Naturwissenschaftler, Mediziner und Ingenieure. (Band 3 der mathematischen Reihe der Lehrbücher und Monographien aus dem Gebiete der exakten Wissenschaften.) 4., unveränd. Aufl. Basel und Stuttgart 1964.

LYLE, Philip: Regression Analysis of Production Costs and Factory Operations. Third edition, rev. by Leonhard H.C. TIPPETT. Edinburgh und London 1957.

MATZ, Adolph: Planung und Kontrolle von Kosten und Gewinn. Handbuch der Planungsrechnung. Deutsche Bearbeitung unter Mitwirkung von A. DEYHLE. (Band 7 der Schriftenreihe der AGPLAN, Hrsg. Josef FUCHS, Ph. KREUZER und K. SCHWANTAG.) Wiesbaden 1964.

MEFFERT, Heribert: Betriebswirtschaftliche Kosteninformationen. Ein Beitrag zur Theorie der Kostenrechnung. (Band 4 der Schriftenreihe: Die Betriebswirtschaft in Forschung und Praxis. Hrsg. E. HEINEN). Wiesbaden 1968.

MELLEROWICZ, Konrad: Kosten und Kostenrechnung. Band I: Theorie der Kosten. 4., durchges. Aufl. Berlin 1963.

MELLEROWICZ, Konrad: Kosten und Kostenrechnung. Band II: Verfahren, 1. Teil: Allgemeine Fragen der Kostenrechnung und Betriebsabrechnung. 4., umgearb. Aufl. Berlin 1966.

MELLEROWICZ, Konrad: Kosten und Kostenrechnung. Band II: Verfahren, 2. Teil: Kalkulation und Auswertung der Kostenrechnung und Betriebsabrechnung. 4., völlig umgearb. Aufl. Berlin 1968.

MENGES, Günter: Ökonometrie. (Die Wirtschaftswissenschaften. Reihe B, Volkswirtschaftslehre, Nr. 20, Hrsg. Erich GUTENBERG.) Wiesbaden 1961.

MOEWS, Dieter: Zur Aussagefähigkeit neuerer Kostenrechnungsverfahren. (Band 39 der Betriebswirtschaftlichen Forschungsergebnisse). Berlin 1969.

NIKLAS, H. und MILLER, M.: Korrelationsrechnung und ihre Anwendung auf Statistik, Versuchswesen, Vererbungslehre, Wissenschaft und Technik. Leipzig (1940).

NORDEN, Helmut und WILLE, Friedrich: Der Betriebsabrechnungsbogen. Die Kostenstellenrechnung in der Praxis. 10., aktualis. u. erw. Aufl. Stuttgart 1965.

NOWAK, Paul: Kostenrechnungssysteme in der Industrie. (Band 21 der Veröffentlichungen der SCHMALENBACH-Gesellschaft). 2., erg. Aufl. Köln und Opladen 1961.

OSTLE, Bernard: Statistics in Research. Basic Concepts and Techniques for Research Workers. Second edition. Ames (Iowa, U.S.A.) 1963.

PACK, Ludwig: Die Elastizität der Kosten. Grundlagen einer entscheidungsorientierten Kostentheorie. (Band 1 der Schriften zur theoretischen und angewandten Betriebswirtschaftslehre). Wiesbaden 1966.

PFANZAGL, Johann: Allgemeine Methodenlehre der Statistik II. Höhere Methoden unter besonderer Berücksichtigung der Anwendungen in Naturwissenschaften, Medizin und Technik. 2., verb. Aufl. Berlin 1966.

RALSTON, Anthony und WILF, Herbert S.: Mathematische Methoden für Digitalrechner. Genehmigte Übersetzung der im Verlag John Wiley & Sons, Inc., New York unter dem Titel "Mathematical Methods for Digital Computers" erschienenen englischsprachigen Ausgabe. München und Wien 1967.

ROCKSTUHL, Joachim P.: Untersuchung über Möglichkeiten einer verursachungsgerechten Zuordnung der im betrieblichen Fertigungsablauf entstehenden Kosten, insbesondere der Restgemeinkosten. Diss. Aachen 1961.

RUMMEL, Kurt: Einheitliche Kostenrechnung auf der Grundlage einer vorausgesetzten Proportionalität der Kosten zu betrieblichen Größen. 3. Aufl. Düsseldorf 1949.

SCHELLING, Hermann von: Die wirtschaftlichen Zeitreihen als Problem der Korrelationsrechnung. Mit besonderer Berücksichtigung der "lag correlation". (Heft 11 der Veröffentlichungen der Frankfurter Gesellschaft für Konjunkturforschung. Hrsg. Eugen ALTSCHUL). Bonn 1931.

SCHMALENBACH, Eugen: Kostenrechnung und Preispolitik. 8., erw. u. verb. Aufl., bearbeitet von Richard BAUER. Köln und Opladen 1963.

SCHNEIDER, Dieter: Die wirtschaftliche Nutzungsdauer von Anlagegütern als Bestimmungsgrund der Abschreibungen. (Band 14 der Beiträge zur betriebswirtschaftlichen Forschung). Köln und Opladen 1961.

SCHNETTLER, Albert: Betriebsanalyse. Stuttgart 1958.

SCHÖNFELD, Hanns-Martin:Kostenrechnung. (Betriebswirtschaftliche Studienbücher. Hrsg. Hans SEISCHAB. Reihe IV: Kaufm. Rechnungswesen). 2., um das Kap. F, V. erw. Aufl. Stuttgart 1964.

SMILLIE, K.W.: An Introduction to Regression and Correlation. Toronto, London und New York 1966.

TINBERGEN, Jan: Einführung in die Ökonometrie. Mit einem Geleitwort von Erich SCHNEIDER. Ins Deutsche übertragen von Hermann CHRISTOPH. Vom Verfasser durchges. u. ergänzt. Wien und Stuttgart 1952.

TINTNER, Gerhard: Handbuch der Ökonometrie. (Enzyklopädie der Rechts- und Staatswissenschaft, begr. von F. von LISZT und W. KASKEL, Hrsg. W. KUNKEL, H. PETERS und E. PREISER. Abteilung Staatswissenschaft). Berlin, Göttingen und Heidelberg 1960.

TSCHUPROW, A.A.: Principles of the Mathematical Theory of Correlation. Translated by M. KANTOROWITSCH. London, Edinburgh and Glasgow 1939.

VORMBAUM, Herbert: Kalkulationsarten und Kalkulationsverfahren. Kalkulationslehre. (Betriebswirtschaftliche Studienbücher. Begr. von Hans SEISCHAB. Reihe IV: Kaufm. Rechnungswesen). Stuttgart 1966.

WALLIS, W. Allen und ROBERTS, Harry V.: Methoden der Statistik. Ein neuer Weg zu ihrem Verständnis. Übersetzt von Harald von WALDHEIM. Mit einem Vorwort von Rolf WAGENFÜHR. Freiburg i. Br. (1960).

WILLE, Friedrich: Plan- und Standardkostenrechnung. Leitfaden. (Betriebswirtschaftliche Bibliothek. Hrsg. Wilhelm HASENACK. Reihe A/VII). 2. Aufl. Essen 1963.

WÖHE, Günter: Einführung in die Allgemeine Betriebswirtschaftslehre. 8., unveränd. Aufl. Berlin und Frankfurt a.M. 1968.

YAMANE, Taro: Statistics, an Introductory Analysis. New York, Evanston, London and Tokyo 1964.

2. Artikel und Aufsätze

AMERMAN, Gilbert: The Mathematics of Variance Analysis. "Journal of Accounting Research". 4 (1953) 3, S. 258 ff., 4 (1953) 4, S. 329 ff., 5 (1954) 1, S. 56 ff.

APEL, Hans: Marginal Cost Constancy and its Implications. "The American Economic Review". 38 (1948) 5, S. 870 ff.

BEECH, Donald G.: Experiences of Correlation Analysis. "Applied Statistics. A Journal of the Royal Statistical Society". 2 (1953) 2, S. 73 ff.

BENSTON, George J.: Multiple Regression Analysis of Cost Behavior. "The Accounting Review". 41 (1966) 4, S. 657 ff.

CHAWIN, I. und KORSHEWITSCH, W.: Die Anwendung mathematischer Methoden und elektronischer Rechenmaschinen für die Analyse der Selbstkosten. "Statistische Praxis. Zeitschrift für Statistik und Rechnungswesen". 19 (1964) 4, S. 107 ff.

CHIU, John S. und De COSTER, Don T.: Multiple Product Costing by Multiple Correlation Analysis. "The Accounting Review". 41 (1966) 4, S. 673 ff.

COMISKEY, Eugene E.: Cost Control by Regression Analysis. "The Accounting Review". 41 (1966) 2, S. 235 ff.

DEAN, Joel: Statistical Cost Curves. "Journal of the American Statistical Association". 32 (1937) 197, S. 83 ff.

DEAN, Joel: The Relation of Cost to Output for a Leather Belt Shop. "National Bureau of Economic Research - Technical Paper". No. 2. New York 1941, S. 3 ff.

DEAN, Joel: Statistical Cost Functions of a Hosiery Mill. "Studies in Business Administration". 11 (1941) 4, S. 5 ff.

EGERMAYER, Frantisek: Die Korrelationsanalyse in der ökonomischen Forschung. "Zeitschrift für Messen, Steuern, Regeln". O. J. (1962), S. 435 ff.

EZEKIEL, Mordecai und WYLIE, Kathryn H.: Cost Functions for the Steel Industry. "Journal of the American Statistical Association". 36 (1941) 213, S. 91 ff.

FÖRSTER, Erhard: Zur Anwendung der Korrelations- und Regressionsanalyse in der Ökonomie. "Wissenschaftliche Zeitschrift der Humboldt-Universität zu Berlin". Gesellschafts- und sprachwissenschaftliche Reihe. 13 (1964) 5, S. 559 ff.

FRAAS, Günther: Neuere Aspekte des Kostenrechnungsprozesses. "Statistische Praxis. Zeitschrift für Statistik und Rechnungswesen". 20 (1965) 10, S. 407 ff., 20 (1965) 11, S. 456 ff.

GARSIDE, M.J.: A Short List of Multiple Regression Programmes Available in Britain."Applied Statistics. A Journal of the Royal Statistical Society". 12 (1963) 3, S. 151 ff.

GAYNOR, Edwin W.: Statistical Testing of Relationships Used in Cost Planning and Allocation. "National Association of Cost Accountants (N.A.C.A.) - Bulletin". Section 1. New York, Mai 1957, S. 1114 ff.

GOLDBERGER, Arthur S. und JOCHEMS, D.B.: Note on Stepwise Least Squares. "Journal of the American Statistical Association". 56 (1961) 293, S. 105 ff.

HABERBECK, Hans-Rudolf: Zur Beschreibung der Abhängigkeitsstruktur des Produktionsfaktorverbrauches. "Zeitschrift für Betriebswirtschaft". 38 (1968) 12, S. 905 ff.

HEINEN, Edmund: Reformbedürftige Zuschlagskalkulation. "Zeitschrift für handelswissenschaftliche Forschung". 10 (1958) 1, S. 1 ff.

JENSEN, Robert E.: A Multiple Regression Model for Cost Control - Assumptions and Limitations. "The Accounting Review". 42 (1967) 2, S. 265 ff.

KNÜPPEL, Helmut, STUMPF, Arthur und WIEZORKE, Bernhard: Mathematische Statistik in Eisenhüttenwerken. Teil I: Regressionsanalyse. "Archiv für das Eisenhüttenwesen". 29 (1958) 8, S. 521 ff.

KNUSSMANN, Rainer: ALGOL-Rechenprogramme statistischer Standardverfahren. "Unternehmensforschung - Operations Research". 5 (1961) 2, S. 61 ff.

KUCERA, Jaroslav: Die Regressionsanalyse mit Hilfe kleiner Ziffernrechner. "Statistische Praxis. Zeitschrift für Statistik und Rechnungswesen". 18 (1963) 12, S. 354 ff.

KWANG, Ching-wen und SLAVIN, Albert: The Simple Mathematics of Variance Analysis. "The Accounting Review", 37 (1962) 3, S. 415 ff.

LORIG, Arthur N.: Joint Cost Analysis as an Aid to Management. "The Accounting Review". 30 (1955) 4, S. 634 ff.

LUDA, Gerhard: Programm zur Durchführung von Regressionsanalysen auf elektronischen Datenverarbeitungsanlagen. "Allgemeines Statistisches Archiv". 46 (1962) 2, S. 122 ff.

MALETZ, Josef: Kostenauflösung. "Zeitschrift für handelswissenschaftliche Forschung". 20 (1926) 7, S. 293 ff.

MANSFIELD, Edwin und WEIN, Harold H.: A Regression Control Chart for Costs. "Applied Statistics. A Journal of the Royal Statistical Society". 7 (1958) 1, S. 48 ff.

MATT, Günter: Die schrittweise Regressionsanalyse und ihre Anwendungsmöglichkeit im kaufmännischen Bereich. "Ablauf- und Planungsforschung - Operational Research". 4 (1963) 4, S. 254 ff. (Diese Schrift ist auch erschienen als Sonderdruck der IBM Deutschland, IBM-Form 78132).

MATT, Günter: Bestimmung statistisch gesicherter Koeffizienten bei der exponentiellen Ausgleichung (Exponential Smoothing). "Unternehmensforschung - Operations Research". 10 (1966) 1, S. 15 ff.

MC CLENON, Paul R.: Cost Finding Through Multiple Correlation Analysis. "The Accounting Review". 38 (1963) 3, S. 540 ff.

MOORE, P.G.: Regression as an Analytical Tool. "Applied Statistics. A Journal of the Royal Statistical Society". 11 (1962) 2, S. 106 ff.

MUIR, Andrew: Marginal Cost. "Applied Statistics. A Journal of the Royal Statistical Society". 7 (1958) 2, S. 96 ff.

NEBELUNG, Carl: Statistische Ermittlung der fixen und variablen Kosten in Betrieben der Großserienfertigung. "Zeitschrift für Betriebswirtschaft". 20 (1950) 7, S. 417 ff.

OLIVER, F.R.: A Cross-section Study of Marginal Cost. "Applied Statistics. A Journal of the Royal Statistical Society". 11 (1962) 2, S. 69 ff.

PLAUT, Hans-Georg: Die Grenz-Plankostenrechnung. "Zeitschrift für Betriebswirtschaft". 23 (1953) 6, S. 347 ff.

PLAUT, Hans-Georg: Unternehmenssteuerung mit Hilfe der Voll- und Grenz-Plankostenrechnung. "Zeitschrift für Betriebswirtschaft". 31 (1961) 8, S. 460 ff.

RIEBEL, Paul: Das Rechnen mit Einzelkosten und Deckungsbeiträgen. "Zeitschrift für handelswissenschaftliche Forschung". Neue Folge. 11 (1959) 3, S. 213 ff.

RUGGLES, Richard: The Concept of Linear Total Cost-Output Regressions. "The American Economic Review". 31 (1941) 2, S. 332 ff.

SCHNEIDER, Dieter: Kostentheorie und verursachungsgemäße Kostenrechnung. "Zeitschrift für handelswissenschaftliche Forschung". Neue Folge. 13 (1961) 11, S. 677 ff.

SCHÖNFELD, Hanns-Martin: Die Anwendung statistischer Methoden in der Kostenrechnung (I). "Kostenrechnungspraxis. Zeitschrift für Betriebsabrechnung, Kostenrechnung und -planung". O.J. (1968) 6, S. 245 ff.

SCHÖNFELD, Hanns-Martin: Die Anwendung statistischer Methoden in der Kostenrechnung (II). "Kostenrechnungspraxis. Zeitschrift für Betriebsabrechnung, Kostenrechnung und -planung". O.J. (1969) 1, S. 19 ff.

SMITH, Caleb A.: The Cost-Output Relation for the U.S. Steel Corporation. "The Review of Economic Statistics". 24 (1942) 4, S. 166 ff.

STAEHLE, Hans: The Measurement of Statistical Cost Functions: An Appraisal of some Recent Contributions. "The American Economic Review". 32 (1942) 2, S. 321 ff.

STEFFEN, Manfred und STEINECKE, Volkmar: Einflußgrößenrechnung zur Kostenplanung eines kontinuierlichen Feinstahlwalzwerkes mit Matrizen. "Stahl und Eisen". 82 (1962) 3, S. 155 ff.

STRAUS, Everet: Cost Accounting and Statistical Cost Functions. "The American Economic Review". 35 (1945) 3, S. 430 ff.

TROXEL, R.B.: Variable Budgets Through Correlation Analysis: A Simplified Approach. "National Association of Accountants (N.A.A.) - Bulletin". New York (1965). S. 48 ff.

WALTERS, A.A.: Expectations and the Regression Fallacy in Estimating Cost Functions. "The Review of Economics and Statistics". 42 (1960) 2, S. 210 ff.

WARTMANN, Rolf: Steuern, Lenken, Planen. Möglichkeiten in einem gemischten Hüttenwerk. "Hoesch-Berichte aus Forschung und Entwicklung unserer Werke". O.J. (1966) 2, 20 ff.

WARTMANN, Rolf, KOPINECK, Hermann-Josef und HANISCH, Wilhelm: Funktionales Arbeiten in Kostenrechnung und Planung mit Hilfe von Matrizen. "Archiv für das Eisenhüttenwesen". 31 (1960) 7, S. 441 ff.

WEBER, Karl: Die Aussagefähigkeit empirischer Kostenfunktionen in betriebswirtschaftlicher Sicht. "Zeitschrift für Betriebswirtschaft". 37 (1967) 1, 1. Ergänzungsheft. S. 47 ff.

WEBER, Karl: Besonderheiten der amerikanischen Kostenrechnung. "Zeitschrift für Betriebswirtschaft". 38 (1968) 11, S. 837 ff.

WYLIE, Kathryn H. und EZEKIEL, Mordecai: The Cost Curve for Steel Production. "The Journal of Political Economy". 48 (1940) 6, S. 777 ff.

YULE, G.U.: Why Do We Sometimes Get Nonsense-Correlations between Time-Series? - A Study in Sampling and the Nature of Time-Series. "Journal of the Royal Statistical Society". 89 (1926) 1, S. 1 ff.

ZANNETOS, Zenon S.: Standard Costs as a First Step to Probabilistic Control: A Theoretical Justification, an Extension and Implications. "The Accounting Review", 39 (1964) 2, S. 296 ff.

3. Broschüren und Informationsschriften

Statistische Programmsammlung für die "olivetti programma 101". Hrsg. von der Deutschen Olivetti GmbH. Direktion Marketing. Frankfurt a.M. Juli 1968.

IBM Application Program. System/360 Scientific Subroutine Package. (360 A - CM - 03 X). Version III. Programmer's Manual. Fourth edition. Hrsg. von International Business Machines Corporation. 1966, 1967, 1968. Nr. H 20 - 0205 - 3.

Korrelationsanalyse IR EIKO. IBM Rechenzentren - Mathematische Statistik. Hrsg. von IBM Deutschland. Mai 1968. IBM-Form 72137-0.

Multiple Regressionsanalyse IR MURA. IBM Rechenzentren - Mathematische Statistik. Hrsg. von IBM Deutschland. Mai 1968. IBM-Form 72140-0.

Siemens System 4004: Statistik-System, Allgemeine Beschreibung. Hrsg. von Siemens AG, o.O. vorl. Ausgabe Oktober 1966. Best.-Nr. 2 - 2605 - 300.

Siemens System 4004: Siemens - Statistik - System SIESTA. Hrsg. von Siemens AG, o.O., Mai 1968. Kennzahl: 543 12 1130 1.

Multiple Linear Regression Program. GE - 400 Series. Application Manual. Hrsg. von Bull General Electric. O.O., Oktober 1965, durchges. Aufl. September 1969. CPB - 1158 C. Ref. 18.35.046 A.

Scientific and technical software: Statistical, Part 1: Statistical analysis, system mark 1. ICL 1900 Series. Hrsg. vom Technical Publications Service der ICL (International Computers Limited). Technical Publication 3345/1. First edition January 1966, reprinted March 1968.

Multiple schrittweise Regressionsanalyse. Hrsg. von Zuse KG., Bad Hersfeld. O.J. Reg.-Nr. 2126.

Programmsystem für automatische Auswahlmethode bei Regressionsanalyse, (modifizierbarer GAUSS-JORDAN- Algorithmus) der ZUSE Z 23. Hrsg. von Zuse KG., Bad Hersfeld, November 1964. Reg.-Nr. 1457.

Statistik-Programme. Programm-Information PI-4-2 des Deutschen Rechenzentrums. Verf. Dr. Friedrich GEBHARDT. Darmstadt, Januar 1968.

Statistik-Programme. Nachträge zu PI-4-2. Programm-Information PI-4-3 des Deutschen Rechenzentrums. Verf. Dr. Friedrich GEBHARDT. Darmstadt, Oktober 1969.

Statistische Programme des Deutschen Rechenzentrums. Teil A: Allgemeines. Programm-Information PI-32 des Deutschen Rechenzentrums. Verf. Dr. Friedrich GEBHARDT. Darmstadt, August 1969.

Statistische Programme des Deutschen Rechenzentrums. Teil B: Einzelbeschreibungen. Programm-Information PI-33 des Deutschen Rechenzentrums. Verf. Dr. Friedrich GEBHARDT. Darmstadt, August 1969.

4. Tabellensammlung

BLEYMÜLLER, Josef: Statistische Tabellen. Zusammenstellung der wichtigsten theoretischen Verteilungen. Münster, Wintersemester 1968/69.

L E B E N S L A U F

Am 1. April 1943 wurde ich als erstes Kind des kaufmännischen Angestellten Emil RECKSIEGEL und seiner Ehefrau Gerda, geb. KOPPMANN, in Bielefeld geboren. Ich besitze die deutsche Staatsangehörigkeit.

Von Ostern 1949 bis Ostern 1953 besuchte ich in Senne I, Kreis Bielefeld, die Volksschule und wechselte anschließend zum Ratsgymnasium Bielefeld über, an dem ich am 12. Februar 1962 die Reifeprüfung ablegte.

Im März 1962 begann ich bei der Baugesellschaft Sudbrack mbH. in Bielefeld eine Lehre als Industriekaufmann, die ich am 31. März 1965 durch Ablegung der Kaufmannsgehilfenprüfung vor der Ostwestfälischen Industrie- und Handelskammer beendete.

Im Sommersemester 1963 hatte ich bereits mein betriebswirtschaftliches Studium an der Westfälischen Wilhelms-Universität Münster aufgenommen und bestand hier am 6. April 1968 die Prüfung als Diplom-Kaufmann.

Nach Beendigung einer einjährigen Tätigkeit als wissenschaftliche Hilfskraft am Institut für Ökonometrie und Wirtschaftsstatistik der Universität Münster bin ich seit Herbst 1969 in einem Industrieunternehmen als Assistent der Geschäftsleitung tätig.